

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 15:55:36 ; Search time 8.12473 Seconds  
(without alignments)  
3199.272 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_585\_2156  
Perfect score: 2848  
Sequence: 1 ATGGGACAGCGGTGACGAC.....CTCTAGAGAGATAAAGAT 1572

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/abss/ABSSWEB\_spool/US10723552/runat\_14022006\_125146\_13079/app\_query.fasta\_1  
-DB=Issued Patents AA -QMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abss02p -USER=US10723552 @CGN 1.1 161 @runat\_14022006\_125146\_13079  
-NCPU=6 -ICPU=3 -NO WMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5 COMB.pcp:\*  
2: /cgn2\_6/ptodata/1/iaa/6 COMB.pcp:\*  
3: /cgn2\_6/ptodata/1/iaa/H COMB.pcp:\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pcp:\*  
5: /cgn2\_6/ptodata/1/iaa/RE COMB.pcp:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2640	92.7	524	2	US-09-171-553B-4
2	1733	60.8	522	2	US-09-075-272-3
3	1733	60.8	622	2	US-09-075-272-2
4	1380.5	48.5	538	2	US-09-309-572-12
5	1380.5	48.5	538	2	US-09-718-096-12
6	1380.5	48.5	1737	2	US-09-309-572-13
7	1380.5	48.5	1737	2	US-09-718-096-13
8	1380	48.5	1312	2	US-09-554-572-26
9	1379.5	48.4	538	2	US-09-370-368-9
10	403.5	14.2	140	2	US-08-850-961-12
11	403.5	14.2	140	2	US-09-479-776-12
12	403.5	14.2	141	2	US-08-850-961-10

13	403.5	14.2	141	2	US-09-479-776-10	Sequence 10, Appl
14	277	9.7	420	2	US-09-719-554-58	Sequence 58, Appl
15	275.5	9.7	378	2	US-08-979-847B-122	Sequence 122, Appl
16	275.5	9.7	398	2	US-08-979-847B-121	Sequence 121, Appl
17	273	9.6	360	2	US-09-719-554-36	Sequence 36, Appl
18	269	9.4	87	2	US-09-938-806A-7	Sequence 7, Appl
19	267.5	9.4	352	2	US-08-979-847B-118	Sequence 118, Appl
20	234	8.2	121	2	US-08-991-789A-2	Sequence 2, Appl
21	234	8.2	121	2	US-09-062-451-2	Sequence 2, Appl
22	234	8.2	121	2	US-09-598-326-2	Sequence 2, Appl
23	234	8.2	121	2	US-09-289-198-2	Sequence 2, Appl
24	234	8.2	121	2	US-09-429-755-2	Sequence 2, Appl
25	234	8.2	121	2	US-09-699-295-2	Sequence 2, Appl
26	234	8.2	121	2	US-09-534-825A-2	Sequence 2, Appl
27	195	6.8	142	2	US-09-719-554-33	Sequence 33, Appl
28	170.5	6.0	779	2	US-10-164-595-56	Sequence 56, Appl
29	163.5	5.7	416	2	US-09-690-454-136	Sequence 136, Appl
30	162.5	5.7	623	2	US-10-104-047-3378	Sequence 3378, Ap
31	161.5	5.7	1185	2	US-09-041-886-23	Sequence 23, Appl
32	161.5	5.7	1185	2	US-09-538-092-1209	Sequence 1209, Ap
33	161	5.7	843	2	US-10-164-595-54	Sequence 54, Appl
34	161	5.7	1184	2	US-09-266-225D-18	Sequence 18, Appl
35	160.5	5.6	553	2	US-09-949-016-7961	Sequence 7961, Ap
36	159	5.6	1344	2	US-09-949-016-10925	Sequence 10925, A
37	157	5.5	1341	2	US-09-949-016-6890	Sequence 6890, Ap
38	154.5	5.4	420	2	US-09-902-540-13993	Sequence 13993, A
39	154	5.4	432	4	PCT-US95-04910-13	Sequence 13, Appl
40	154	5.4	567	2	US-09-205-258-573	Sequence 573, Appl
41	154	5.4	567	2	US-10-004-860-573	Sequence 573, Appl
42	154	5.4	707	2	US-09-919-039-278	Sequence 278, Appl
43	154	5.4	707	2	US-09-538-092-993	Sequence 993, Appl
44	150	5.3	719	2	US-09-949-016-7766	Sequence 7766, Ap
45	150	5.3	919	2	US-09-949-016-6954	Sequence 6954, Ap

ALIGNMENTS

RESULT 1  
US-09-171-553B-4  
; Sequence 4, Application US/09171553B  
; Patent No. 6756227  
; GENERAL INFORMATION:  
; APPLICANT: GALBRAITH, DANIEL N.  
; APPLICANT: HAMWORTH, CHRISTINE  
; APPLICANT: LEE, GILLIAN M.  
; APPLICANT: SMITH, KENNETH T.  
; TITLE OF INVENTION: PORCINE RETROVIRUS  
; FILE REFERENCE: CFV-5.01  
; CURRENT APPLICATION NUMBER: US/09/171,553B  
; CURRENT FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: PCT/GB97/01087  
; PRIOR FILING DATE: 1997-04-18  
; PRIOR APPLICATION NUMBER: GB 9702668.6  
; PRIOR FILING DATE: 1997-02-10  
; PRIOR APPLICATION NUMBER: GB 9608164.1  
; PRIOR FILING DATE: 1996-04-19  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Porcine retrovirus  
US-09-171-553B-4

Alignment Scores:  
Pred. No.: 3.64e-236 Length: 524  
Score: 2640.00 Matches: 495  
Percent Similarity: 97.3% Conservative: 16  
Best Local Similarity: 94.3% Mismatches: 13  
Query Match: 92.7% Indels: 2  
DB: 0

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-09-171-553B-4 (1-524)

QY 1 ATGGGACAGCGGTGACGACCCCTCTTAGTTTGTACTCTCGACCATTTGGACTGAAGTTAAA 60  
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20  
QY 61 TCCAGGGCTCATAAATTGCTCAGTTTAAAGAGGACCTTGGCAGACTTTCTGTGTC 120  
Db 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysAla 40  
QY 121 TCTGAATGGCCGACATTCGATGTGGATGGCCATCAGAGGGGACCTTTAAATTCGAGATT 180  
Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
QY 181 ATCTTGGCTGTAAAGCAGTTATTTTCAGACTGGACCCGGCTCATCCCGATCAGGAG 240  
Db 61 IleLeuAlaValLysAlaIleIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
QY 241 CCCTATATCTTACGTGGCAAGATTGGCAGAGGATCTCCGCCATGGTGTAAACCATGG 300  
Db 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100  
QY 301 CTGAATAAGCCAGAAACCGAGTCCCGAATTCTGGCTCTTGGAGAGAAAAACAACAC 360  
Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
QY 361 TCGGCTGAAAAGTCAAGCCCTC-TCCTCATATCTACCCGAGATTGAGGAGCCACCGGC 419  
Db 121 SerAlaGluLysValGluProSerSerSerTyrLeuProArgPargGlyAlaAlaAsp 140  
QY 420 TTGGCCGGAACCCCAATCTGTTCCCCACCCCTTATCTGGCACAGGTCGCCGAGGGG 479  
Db 141 LeuAlaGlyThrProThrCysSerProThrProLeuSerSerThrGlyCysGluGly 160  
QY 480 ACCCTTTGCCCTCTCTGAGTTCGGCGTGGAGGACCTGCTGCGAGGACTCGAGCCG 539  
Db 161 Thr-SerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerAr 180  
QY 540 GAGGGGCCACCCCGGAGCGGACAGAGAGATCGGCATTTACCGTCGCCACGTAGCG 599  
Db 180 GArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGI 200  
QY 600 CCTCCACACCGGGGGCCAAATTCAGCCCTCCAGTATTGGCCCTTTCTTCGAGA 659  
Db 200 YProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAs 220  
QY 660 TCTCTAATTTGAAAACTAACCATCCCTTTTCTGGAGGATCCCAACGCTCACGGG 719  
Db 220 pLeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGI 240  
QY 720 GTTGTGGTGGTCCCTTATGTTCTCTCACCAGCTACTTGGGATGATTCTCACAGCTGCT 779  
Db 240 YLeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLe 260  
QY 780 CGAGACACTCTTCAACCGGAGGCGGACAGAGAATTTCTATAGAGCTAGAAAAAATGT 839  
Db 260 uGlnThrLeuPheThrThrGluLargLargLargIleLeuLeuGluAlaArgLysAsnVa 280  
QY 840 TCCTGGGCGCGAGCGGACCCACCGGTTGCAAAATGAGATTGATGGGATTTCCCTT 899  
Db 280 lProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLe 300  
QY 900 AACTCGCCCGTGGGACTCAACACCGCTGAAGGTAGGAGACTTGAATAATCTATCG 959  
Db 300 uThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrAr 320  
QY 960 CCAGGCTCTGGTGGGGTCTCCGGGGCGCTCAAGACGGCCACTAATTTGGCTTAAGT 1019  
Db 320 gGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVa 340  
QY 1020 AAGAGAAGTGTACGAGGACCGAATGAACCCCTCTGTTTCTTGAGAGGCTCTTGA 1079  
Db 340 lArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGI 360

QY 1080 AGCCTTCAGGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAGCCTCAGTGGC 1139  
Db 360 uAlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAl 380  
QY 1140 TTTGGCCTTTATAGGACAGTCAGCTTGGATATTAGAAAAGCTTCAGACACTGGAAG 1199  
Db 380 aLeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGI 400  
QY 1200 GTTACAGGCGCTCAGTTTACGTGATCTAGTGAAGGAGCAGAGAAAGTATATTACAAAAG 1259  
Db 400 YLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgAr 420  
QY 1260 GGACACAGAAAGAAAGGGAACAAAGAAAAGAGAGAGAAAGAGAGGAGGAGGAAAG 1319  
Db 420 gGluThrGluGluGluLysGluGlnArgLysGluLysGluArgGluGluArgGluGluAr 440  
QY 1320 ACCTAATAACCGCAAGAGAAATTTGACTAAGATCTTGGCTGCTGAGTGGTTGAAGGAA 1379  
Db 440 gArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLy 460  
QY 1380 AAGCAATACGAAAGAGAGAGAGATTTTAGGAAAATTAGGTACGCCCTCAGACAGTCAAG 1439  
Db 460 sSerSerArgGlnArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGI 480  
QY 1440 GAACCTGGCAATAGGACCCCACTCGACAAGGACCAATGTGCATATTGTAAAGAAAGAGG 1499  
Db 480 YAsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGI 500  
QY 1500 ACCTGGGCAAGAACTGCCCAAGAGGGGAAAACAAAGGACCAAGGATCCTAGCTCTAGA 1559  
Db 500 YHisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGI 520  
QY 1560 AGAAGATAAAGAT 1572  
Db 520 uGluAspLysAsp 524  
RESULT 2  
US-09-075-272-3  
; Sequence 3, Application US/09075272  
; Patent No. 6136598  
; GENERAL INFORMATION:  
; APPLICANT: MILLER, A. DUSTY  
; APPLICANT: WOLGAMOT, GREG  
; APPLICANT: BONHAM, LYNN  
; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL  
; TITLE OF INVENTION: PACKAGING CELL LINES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/075,272  
; FILING DATE: 08-MAY-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/046,140  
; FILING DATE: 09-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pooh, Brian W.  
; REGISTRATION NUMBER: 32,928  
; REFERENCE/DOCKET NUMBER: 14538A-003710  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 467-9600  
; TELEFAX: (415) 576-0300



## ; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 522 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-09-075-272-3

## Alignment Scores:

Pred. No.: 5,1e-152 Length: 522  
 Score: 1733.00 Matches: 340  
 Percent Similarity: 73.3% Conservative: 61  
 Best Local Similarity: 62.2% Mismatches: 96  
 Query Match: 60.8% Indels: 50  
 DB: 2 Gaps: 12

US-10-723-552-3\_copy\_585\_2156 (1-1572) x US-09-075-272-3 (1-522)

```

QY 1 ATGGGACAGCGGTGACGACCCCTCTAGTTTGACTCTGACCATTTGGACTGAAGTTAAA 60
Db 1 MetGlyGlnSerValSerThrProLeuSerLeuThrLeuGluHisTrpLysGluValys 20
QY 61 TCCAGGCTCATATTTGTCAGTTTCAGTTTAAGAGGACCTTGCAGACTTCTGTGTC 120
Db 21 IleArgAlaHisAsnGlnSerValGluValArgLysGlyProTrpGlnThrPheCysAla 40
QY 121 TCTGAATGGCGGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTTCTGAGATT 180
Db 41 SerGluTrpProThrPheGlyValGlyTrpProGluGlyAlaPheAspLeuSerLeu 60
QY 181 ATCTGCTGTTAAAGCAGTATTTTCAGACTGCGACCGGCTCTCATCCGATCCGAG 240
Db 61 IleAlaAlaValArgGlyIleValPheGlnGlu---GluGlyGlyHisProAspGlnIle 79
QY 241 CCCTATATCTTACGTGCGAAGTTTTCGAGAGGATCTCGCCATCGSTTAACCATGG 300
Db 80 ProTyrIleValThrTrpGlnSerLeuValGlnPheProSerTrpValLysProTrp 99
QY 301 CTGAATAAGCCACAGAAAGCCAGGTCCCGCAATTTCTGGCTCTTGGAGAGAAAAACAAAC 360
Db 100 -----ThrProAsnProSerLysLeuThrValAlaValAlaGlnSerAsp 114
QY 361 TCGGCTGAAAAGTCAAGCCCTCT-----CCTCATATCTACCCCGAGATTGAG----- 408
Db 115 AlaAlaGluLysSerGlyProSerAlaProProLysIleTyrProGluIleAspLeu 134
QY 409 -----GAGCCACCGCTTGGCGGAAACCCCAATCTGTCCC-----CCA 447
Db 135 LeuTrpMetAspSerGlnProProTyrProLeuProGlnGlnProProAlaAlaAla 154
QY 448 CCCCCTTATCTGGCACAGGGTGGCGGAGGGGACCTTTGCCCCCTCTCGAGGCTCCGCG 507
Db 155 ProProValAlaAla-----ProGlnProGluProThrAlaSerGly 168
QY 508 GTGAGGGACCTGCTCGAGGACTCGAGCGGAGGGGCGCCACCCCGAGCGG----- 561
Db 169 AlaGlnGlyProAlaGlyIleThrArgSerArgGlyArgSerProAlaGluGluGly 188
QY 562 ---ACACACAGATCGCGACATTACCGCTCGCAGCGTAC---GGCCCTCCACACCGGG 615
Db 189 GlyProAspSerThrValAlaLeuProLeuArgAlaHisValGlyGlyProThrProGly 208
QY 616 ---GGCCAAATTGACGCCCTCCAGTATTGGCCCTTTTCTCTGAGATCTCTATAATTGG 672
Db 209 ProAsnAspLeuLeuProLeuGlnTyrTrpProPheSerSerSerAspLeuTyrAsnTrp 228
QY 673 AAACTAACCATCCCTTTCTCGAGAGATCCCAACGCTCAAGGGGTTGGTGGAGTCC 732
Db 229 LysThrAsnHisProProPheSerGluAsnProSerGlyLeuThrGlyLeuLeuGluSer 248
QY 733 CTTATGTTCTCTACACGCTTCTGGATGATTGTTCAACAGCTGTGCGACACTCTTC 792
Db 249 LeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnValLeuPhe 268

```

## RESULT 3

US-09-075-272-2

; Sequence 2, Application US/09075272  
 ; Patent No. 6136598  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MILLER, A. DUSTY  
 ; APPLICANT: WOLGAMOT, GREG  
 ; APPLICANT: BOHAM, LYNN  
 ; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL  
 ; TITLE OF INVENTION: PACKAGING CELL LINES  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
 ; STREET: Two Embarcadero Center, 8th Floor  
 ; CITY: San Francisco  
 ; STATE: California

```

QY 793 ACAACCGAGGACGAGAGAAATTTCTATTAGAGCTAGAAAAAATGTTCTCTGGGCCGAC 852
Db 269 ThrThrGluGluArgGluArgIleLeuMetGluAlaArgLysAsnValLeuGlyGluAsp 288
QY 853 GGGCGACCCACGCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTAACTCGCCCCCGT 912
Db 289 GlyThrProThrAlaLeuProAsnLeuValAspGluAlaPheProLeuAsnArgProAsn 308
QY 913 TGGGACTACAAACACGCTGAGGTAGGAGAGCTTGAATAATCTATCCGACGCTCTGGTG 972
Db 309 TrpAspTyrAsnThrAlaGluGlyArgGlyArgLeuLeuValTyrArgThrLeuVal 328
QY 973 GCGGGTCTCCGGGCGGCTCAAGACGCGCCCACTAATTTGGCTAAGTAAAGAGAAGTGATG 1032
Db 329 AlaGlyLeuArgGlyAlaAlaArgProThrAsnLeuAlaValArgGluValLeu 348
QY 1033 CAGGGACCGAATGAACCCCTCTGTTTCTTCTGAGAGGCTCTTGAAGACCTTCAAGCGG 1092
Db 349 GlnGlyGlnThrGluProProSerValPheLeuGluArgLeuMetGluAlaTyrArgArg 368
QY 1093 TACACCCCTTTTGTATCCACCTCAGAGCGCCCAAAAGCCTCAGTGGCTTTCGSCCTTTATA 1152
Db 369 TyrThrProPheAspProSerSerGlyGlyGlnLysAlaValAlaMetAlaPheIle 388
QY 1153 GGACAGTCAGCCTTTGGATATTAGAAAGAGCTTCAGAGACTTGAAGGGTTACAGAGGCT 1212
Db 389 GlyGlnSerAlaProAspIleLysLysLeuGlnArgLeuGluGlyLeuGlnAspTyr 408
QY 1213 GAGTTAGTCTAGTCTAGTGAAGGAGGAGAGAAAGTATATTACAAAAGGAGACAGAGAA 1272
Db 409 ThrLeuGlnAspLeuValLysGluAlaGluLysValTyrHisLysArgGluThrGluGlu 428
QY 1273 GAAAGGAAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1332
Db 429 GluArgGlnGluArgGluLysLysGluValGluGluArgGluAsnArgAspArgArg 448
QY 1333 CAAGAGAGAAATTTGACTTAAGATCTTGGCTGCGCTGCTTGAAGGAGAGAGAGAGAGAG 1392
Db 449 GlnGluArgAsnLeuSerLysIleLeuAlaValIleAsnAspArg----- 464
QY 1393 AGAGAGAGATTTTAGGAAAAATTAGTCTAGCGCCCTAGACAGTCAAGGAACTCGGGCAAT 1452
Db 465 -----GlnSerGluLysGlyArgThrGlyPheLeuGlyAsn 476
QY 1453 AGG-----ACCCCACTCGACAGAGGACCATGTGCATAT 1485
Db 477 ArgAlaValLysProGlyGlyArgLysThrProLeuGluLysAspGlnCysAlaPhe 496
QY 1486 TGTAAGAGAGAGAGACACTGGCAAGAACTGCCCCAAGAGAGAGAGAGAGAGAGAGAGAG 1545
Db 497 CysLysGluLysGlyHisTrpAlaLysAspCysProLysLys---ArgArgGlnPheLys 515
QY 1546 ATCCTAGCTCTAGAAGAGAT 1566
Db 516 ValLeuThrLeuGluAspAsp 522

```







Alignment Scores:		
Pred. No.:	4.36e-119	Length: 1737
Score:	1380.50	Matches: 283
Percent Similarity:	62.2%	Conservative: 69
Best Local Similarity:	50.0%	Mismatches: 141
Query Match:	48.5%	Indels: 73
DB:	2	Gaps: 13
US-10-723-552-3_COPY_585_2156 (1-1572) x US-09-309-572-13 (1-1737)		
Qy	1	ATGGGACAGCGGTGACGACCCCTCTTAGTTTGACTCTCGACCATCGGACTGAAGTTAAA 60
Db	1	MeGlyGlnThrValThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20
Qy	61	TCCAGGCGCTCATAAATTTGTCCAGTTTCAGTTTAAGAAGGGACCTTGGCAGACTTCTCTGTGC 120
Db	21	ArgileAlaHisAsnGlnSerValAspValLysLysArgArgTrpValThrPheCysSer 40
Qy	121	TCGTGAATGGCGGACATTCGATGTGTGATGCCCATCAGAGGGACCTTTAATCTCGAGATT 180
Db	41	AlaGlnTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60
Qy	181	ATCTCGGCTCTTAAAGCAGTATTATTTTCAGACTGGACCGGCTCTCATCCGATCAGGAG 240
Db	61	IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80
Qy	241	CCCTATATCTTACGTGGCAAGATTTTGGCAGAGGATCTCCCGCATGGGTAAACCATGG 300
Db	81	ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProTrpValLysProPhe 100
Qy	301	CTGAATTAAGCAAGAAGCCAGGTGCCCGAATTCTTGCTCTTGGAGAGAAAAACAACAC 360
Db	101	ValHis---ProLysProProProProLeuProProSerAlaProSerLeuProLeuGlu 119
Qy	361	TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTACCCGAGATTGGAGGCCA---CCG 417
Db	120	ProProArgSerThrProProArgSerSerLeuThrProAlaLeuThrProSerLeuGly 139



Qy	1324	AATAAACGGCAAGAGAGAATTTCAGCTAAGATCTTGGCTGCAGTGGTTGAAGCGAAAGC	1383
Db	467	AspArgArgHisArgGluMetSerLysLeuLeuAlaThrValValSerGlyGlnLys	486
Qy	1384	AATACGGAAAGAGAGAGAGATTTTATAGGAAATATTAGTTCAGGCCCTAGACAGTCAGGGAAC	1443
Db	487	Gln-----AspArgGlnGlyGlyGlu	493
Qy	1444	CTGGGCAATATAGGACCCCACTCGACAAGGACCAATGTGCATATTGTAAAGAAAGAGAGACAC	1503
Db	494	--ArgArgArgSerGlnLeuAspArgAspGlnCysalaTyCysLysGluLysGlyHis	512
Qy	1504	TGGGCAAGGAACCTGCCCCAAGAAG-----GGAACAAGAGCAACGAGG-----	1545
Db	513	TrpAlaLysAspCysProLysLysProArgGlyProArgGlyProArgProGlnThrSer	532
Qy	1546	ATCCTGACTCTAGAGAAGAA	1563
Db	533	LeuLeuThrLeuAspAsp	538
RESULT 8			
US-09-554-572-26			
; Sequence 26, Application US/09554572			
; Patent No. 6573091			
; GENERAL INFORMATION:			
; APPLICANT: NATURE TECHNOLOGY, INC.			
; TITLE OF INVENTION: CHIMERIC VIRAL PACKAGING SIGNAL WITHOUT GAG GENE			
; TITLE OF INVENTION: SEQUENCES			
; FILE REFERENCE: 228.00030201			
; CURRENT APPLICATION NUMBER: US/09/554,572			
; CURRENT FILING DATE: 2000-09-18			
; NUMBER OF SEQ ID NOS: 26			
; SOFTWARE: Patentin ver. 2.0			
; SEQ ID NO 26			
; LENGTH: 1312			
; TYPE: PRT			
; ORGANISM: Murine leukemia virus			
US-09-554-572-26			
Alignment Scores:			
Pred. No.: 4,29e-119 Length: 1312			
Score: 1380.00 Matches: 283			
Percent Similarity: 62.0% Conservative: 68			
Best Local Similarity: 50.0% Mismatches: 141			
Query Match: 48.5% Indels: 74			
DB: 2 Gaps: 13			
US-10-723-552-3_COPY_585_2156 (1-1572) x US-09-554-572-26 (1-1312)			
Qy	1	ATGGGACAGAGCGTGGAGGACCCCTCTAGTTTGCATCTCGACCATTTGGACTGAAGTTAAA	60
Db	111	MetGlyGlnThrValThrThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu	130
Qy	61	TCCAGGCGCTCATAAATTGTCAGTTTCAGGTTTAAAGAGGGACCTTGGCAGACTTTCTCTGTC	120
Db	131	ArgIleAlaHisGlnSerValAspValLysLysArgArgTrpValThrPheCysSer	150
Qy	121	TCTGAATGCGCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATCTGAGATT	180
Db	151	AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu	170
Qy	181	ATCCTGGCTGTTAAAGCAGTATTATTTTCAGACTGGGACCCGGCTCTCATCCGATCAGGAG-	240
Db	171	IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal	190
Qy	241	CCCTATATCTTACGTGGCAAGATTTTGGCAGAGGATCTCCCGCATGGGTAAACCATGG	300
Db	191	ProTyrlleValThrTrpGluAlaLeuAlaPheAspProProTrpValLysProPhe	210
Qy	301	CTGAATTAAGCCAAAGACCGAGTCCCGAATTTCTGGCTTTGGAGAGAAAAACAACAC	360
Db	211	ValHis---ProLysProProProLeuProProSerAlaProSerLeuProLeuGlu	229







```

Db 455 ArgArgThrGluAspGluGlnLysGluArgAspArgArgHisArgGluMet 474
QY 1348 ACTAAGATCTTGGCTGCAAGTGTTCGAAGGAAAGCAATACGAAAGAGAGAGATTTT 1407
Db 475 SerArgLeuLeuAlaThrValSerGlyGlnArgGlnAspArgGlnGluGluArg 494
QY 1408 AGGAAATATTAGTCAAGCCCTAGACAGTCAGGGAACCTGGGCAATAGGACCCCACTCGAC 1467
Db 495 Arg-----ArgSerGlnLeuAsp 500
QY 1468 AAGGACCAATGTCATATTGTAAGAAAGAGAGACACTGGGCAAGAACTGCCCAAG--- 1524
Db 501 CysAspGlnCysThrTyrcysGluGlnGlnGlyHisTrpAlaLysAspCysProArgArg 520
QY 1525 ---AAGGAAACAAAGGACCAAGG-----ATCCTAGTCTCTAGAAGAA 1563
Db 521 ProArgGlyProArgGlyProArgProGlnThrSerLeuLeuThrLeuAspAsp 538

RESULT 10
US-08-850-961-12
; Sequence 12, Application US/08850961
; Patent No. 6013517
; GENERAL INFORMATION:
; APPLICANT: Respass, James G.
; APPLICANT: De Polo, Nicholas J.
; APPLICANT: Chada, Sunil
; APPLICANT: Sauter, Sybille
; APPLICANT: Bodner, Mordechai
; APPLICANT: Driver, David A.
; TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation, Intellectual Property - R440
; STREET: P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,961
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kruse, No. 6013517man J.
; REGISTRATION NUMBER: 35,235
; REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-3520
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-850-961-12

Alignment Scores:
Pred. No.: 6.37e-29 Length: 140
Score: 403.50 Matches: 76
Percent Similarity: 61.1% Conservative: 15
Best Local Similarity: 51.0% Mismatches: 35
Query Match: 14.2% Indels: 23
DB: 2 Gaps: 2

US-10-723-552-3_COPY_585_2156 (1-1572) x US-08-850-961-12 (1-140)

```

```

QY 1 ATGGGACAGACGGTCAGCACCCCTCTTAGTTTGGACTCTCGACCAATGTCGACTGAAGTTAAA 60
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20
QY 61 TCCAGGGCTCATATTTGTCTAGTTCAAGTTAAGAGGACCTTGGCAGACTTTCTGTGTC 120
Db 21 ArgIleAlaHisAsnGlnSerValAspValLysLysArgArgTrpValThrPheCysSer 40
QY 121 TCTCAATGGCCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCAGATT 180
Db 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60
QY 181 ATCTGGCTGTATAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 240
Db 61 IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80
QY 241 CCTATATCTCTAGTCGCAAGATTTGGCAGAGAGATCTCGCCATGGTTAAACCATGG 300
Db 81 ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProTrpValLysProPhe 100
QY 301 CTGAATAAGCCAGAAAGCCAGGTCCTCCGAATTCCTGGCTCTTGGAGAGAAAAACAACAC 360
Db 101 ValHisPro----- 103
QY 361 TCGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCGAGATTGAGGAGCCCGGCT 420
Db 104 -----LysProProProLeuProProSerAlaProSerLeuProLeu 118
QY 421 TGGCCGAAACCCCAATCTGTTCCCCCA 447
Db 119 GluPro---ProArgSerThrProPro 126

RESULT 11
US-09-479-776-12
; Sequence 12, Application US/09479776
; Patent No. 633195
; GENERAL INFORMATION:
; APPLICANT: Respass, James G.
; APPLICANT: De Polo, Nicholas J.
; APPLICANT: Chada, Sunil
; APPLICANT: Sauter, Sybille
; APPLICANT: Bodner, Mordechai
; APPLICANT: Driver, David A.
; TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: INTELLECTUAL PROPERTY-R440
; P.O. BOX 8097
; CITY: EMERYVILLE
; STATE: CA
; COUNTRY: USA
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,776
; FILING DATE: 07-Jan-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: KRUSE, NORMAN J.
; REGISTRATION NUMBER: 35,235
; REFERENCE/DOCKET NUMBER: 930049.424C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)622-4900
; TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid

```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-479-776-12

Alignment Scores:
Pred. No.: 6,37e-29 Length: 140
Score: 403.50 Matches: 76
Percent Similarity: 61.1% Conservative: 15
Best Local Similarity: 51.0% Mismatches: 35
Query Match: 14.2% Indels: 23
DB: 2 Gaps: 2

US-10-723-552-3_COPY_585_2156 (1-1572) x US-09-479-776-12 (1-140)
QY 1 ATGGACAGAGCGGTGACGACCCCTTTAGTTGACTCTCGACCAATGAGCTGAAGTTAAA 60
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20
QY 61 TCCAGGGCTCATATTTGTCAGTTTCAGGTTAAGAGGGACCTTGCAGACTTCTGTGTC 120
Db 21 ArgIleAlaHisAsnGlnSerValAspValLysLysArgArgTrpValThrPheCysSer 40
QY 121 TCTGAATGCGCGACATTCGATGTTGAGTTCAGTTTAAATTTCTGAGATT 180
Db 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60
QY 181 ATCTGGCTGTTAAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 240
Db 61 IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80
QY 241 CCCTATATCTTACGTGCGCAAGATTTCGAGAGGATCTCCGCCATGGGTTAAACCAATGG 300
Db 81 ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProProTrpValLysProPhe 100
QY 301 CTGAATAGCCCAAGAACCCAGGTCCCGCAATTTCTGGCTCTTGGAGAGAAAAACAAACAC 360
Db 101 ValHisPro----- 103
QY 361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGAGCCACCGCT 420
Db 104 -----LysProProProProLeuProProSerAlaProSerLeuProLeu 118
QY 421 TGGCGGAACCCCAATCTGTTCCCA 447
Db 119 GluPro----ProArgSerThrProPro 126
```

## RESULT 12

```
; Sequence 10, Application US/08850961
; Patent No. 6013517
; GENERAL INFORMATION:
; APPLICANT: Respass, James G.
; APPLICANT: De Polo, Nicholas J.
; APPLICANT: Chada, Sunil
; APPLICANT: Sauter, Sybille
; APPLICANT: Bodner, Mordechai
; APPLICANT: Driver, David A.
; TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation, Intellectual Property - R440
; STREET: P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/850,961
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kruse, No. 6013517man J.
; REGISTRATION NUMBER: 35,235
; REFERENCE/DOCKET NUMBER: 930049.424C4 / 1147.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-3520
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 141 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-850-961-10

Alignment Scores:
Pred. No.: 6,39e-29 Length: 141
Score: 403.50 Matches: 76
Percent Similarity: 61.1% Conservative: 15
Best Local Similarity: 51.0% Mismatches: 35
Query Match: 14.2% Indels: 23
DB: 2 Gaps: 2

US-10-723-552-3_COPY_585_2156 (1-1572) x US-08-850-961-10 (1-141)
QY 1 ATGGACAGAGCGGTGACGACCCCTTTAGTTGACTCTCGACCAATGAGCTGAAGTTAAA 60
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20
QY 61 TCCAGGGCTCATATTTGTCAGTTTCAGGTTAAGAGGGACCTTGCAGACTTCTGTGTC 120
Db 21 ArgIleAlaHisAsnGlnSerValAspValLysLysArgArgTrpValThrPheCysSer 40
QY 121 TCTGAATGCGCGACATTCGATGTTGAGTTCAGGTTAAATTTCTGAGATT 180
Db 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60
QY 181 ATCTGGCTGTTAAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 240
Db 61 IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80
QY 241 CCCTATATCTTACGTGCGCAAGATTTCGAGAGGATCTCCGCCATGGGTTAAACCAATGG 300
Db 81 ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProProTrpValLysProPhe 100
QY 301 CTGAATAGCCCAAGAACCCAGGTCCCGCAATTTCTGGCTCTTGGAGAGAAAAACAAACAC 360
Db 101 ValHisPro----- 103
QY 361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGAGCCACCGCT 420
Db 104 -----LysProProProProLeuProProSerAlaProSerLeuProLeu 118
QY 421 TGGCGGAACCCCAATCTGTTCCCA 447
Db 119 GluPro----ProArgSerThrProPro 126

RESULT 13
US-09-479-776-10
; Sequence 10, Application US/09479776
; Patent No. 6333195
; GENERAL INFORMATION:
; APPLICANT: Respass, James G.
; APPLICANT: De Polo, Nicholas J.
; APPLICANT: Chada, Sunil
; APPLICANT: Sauter, Sybille
; APPLICANT: Bodner, Mordechai
; APPLICANT: Driver, David A.
; TITLE OF INVENTION: CROSSLESS RETROVIRAL VECTORS
; NUMBER OF SEQUENCES: 45
```

CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHIRON CORPORATION  
STREET: INTELLECTUAL PROPERTY-R440  
P.O. BOX 8097  
CITY: EMERYVILLE  
STATE: CA  
COUNTRY: USA  
ZIP: 94662-8097  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,776  
FILING DATE: 07-Jan-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: KRUSE, NORMAN J.  
REGISTRATION NUMBER: 35,235  
REFERENCE/DOCKET NUMBER: 930049.424C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)622-4900  
TELEFAX: (206)682-6031  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 141 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-479-776-10

Alignment Scores:  
Pred. No.: 6 39e-29 Length: 141  
Score: 403.50 Matches: 76  
Percent Similarity: 61.1% Conservative: 15  
Best Local Similarity: 51.0% Mismatches: 35  
Query Match: 14.2% Indels: 23  
DB: 2 Gaps: 2

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-09-479-776-10 (1-141)

Qy 1 ATGGGACAGCGGTGACGACCCCTCTTAGTTTGACTCTCGACCATTCGAGTGAAGTAA 60  
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20  
Qy 61 TCCAGGGCTCATATTTGTGAGTTCAGTTTAAAGAGGACCTTGGCAGACTTCTGTGTC 120  
Db 21 ArgIleAlaHisAsnGlnSerValAspValLysLysArgTrpValThrPheCysSer 40  
Qy 121 TCTGAATGGCGGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTAAATTCGAGATT 180  
Db 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60  
Qy 181 ATCTCGGTGTTAAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCGATCAGAG 240  
Db 61 IleThrGlnValLysIleLeuValPheSerProGlyProHisGlyHisProAspGlnVal 80  
Qy 241 CCCTATATCTTACGTGCAAGATTGGCAGAGGATCTCCGCCATCGGTTAAACCAAGG 300  
Db 81 ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProTrpValLysProPhe 100  
Qy 301 CTGAATAAGCCAAAGACCCAGGTCCCGAATTCCTGGCTCTTGAGAGAAAAACAACAC 360  
Db 101 ValHisPro----- 103  
Qy 361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCGAGATTGAGAGCCACCGGCT 420  
Db 104 -----LysProProProProProProSerAlaProSerLeuProLeu 118  
Qy 421 TGGCCGGAACCCCAATCTGTTCCCA 447  
||| |||:||||| |||||

Db 119 GluPro---ProArgSerThrProPro 126  
RESULT 14  
US-09-719-554-58  
; Sequence 58, Application US/09719554  
; Patent No. 6919438  
; GENERAL INFORMATION:  
; APPLICANT: ALLIEL, Patrick  
; APPLICANT: PERIN, Jean-Pierre  
; APPLICANT: RIERGER, Francois  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE AND DEDUCED PROTEIN SEQUENCE FAMILY WITH I  
; TITLE OF INVENTION: ENDOGENOUS RETROVIRAL MOTIFS  
; FILE REFERENCE: 200936USOPCT  
; CURRENT APPLICATION NUMBER: US/09/719,554  
; CURRENT FILING DATE: 2001-12-26  
; PRIOR APPLICATION NUMBER: PCT/FR99/01513  
; PRIOR FILING DATE: 1999-06-25  
; NUMBER OF SEQ ID NOS: 122  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 58  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (21)..(21)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (86)..(86)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (166)..(166)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (188)..(188)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (331)..(331)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (361)..(361)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (362)..(362)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (374)..(374)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (380)..(380)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (382)..(382)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (408)..(408)  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc feature  
; LOCATION: (413)..(413)  
; OTHER INFORMATION: Xaa is any amino acid  
US-09-719-554-58

Alignment Scores:  
Pred. No.: 5.6e-17 Length: 420  
Score: 277.00 Matches: 106  
Percent Similarity: 39.6% Conservative: 50  
Best Local Similarity: 26.9% Mismatches: 172  
Query Match: 9.7% Indels: 66  
DB: 2 Gaps: 11

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-09-719-554-58 (1-420)

Qy 517 CCTGCTGACGGACTCGAGCGGGGGCCACCCCGGAGCGGACAGATC--- 573



```

Db      27 GlnThr-----LeuArgLysLysArgPheIlePhePheCysSerThr 40
Qy      124 GAATGCCGACATTCGATGTT-----GGATGCCCATCAGAGGGGACCTTTAAT 171
Db      41 AlaTyrProGlnTyrProLeuGlnGlyArgGluThrTrpLeuProGluGlySerIleAsn 60
Qy      172 TCTGAGATTATCTGGCTGTAAACAGTATTATTTTCAGACTGGACCGGCTCTCATCCC 231
Db      61 TyrAsnIleIleLeuGlnLeuAspLeuPheCysArgLysGluGlyLysTrpSer----- 78
Qy      232 GATCAGGAGCCCTATATCTTACGTGGCAAGATTGGCAGAGGATCCTCCGCCATGGGTT 291
Db      79 ---GluValProTyrValGlnThrPheSerLeuArgAsp----- 91
Qy      292 AAACCATGGCTGAATAAGCAAGAAAGCCAGGTCCCGAATTCCTGGCTCTTGGAGAGAA 351
Db      91 ----- 91
Qy      352 AACAAACACTCGGTGAAAGAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGAG 411
Db      92 AsnSerGlnLeuCysLysLysCysGlyLeuCysProThrGlySerPro-----GlnSer 109
Qy      412 CCACCGGCTTGGCGGACCCCAATCTGTCTCCCCACCCCTTATCTGGCACAGGTTGCC 471
Db      110 ProProTyrPro-----SerValProSerPro----- 119
Qy      472 GCGAGGGGACCCCTTTGCCCTCTCGGAGCTCCGGGTGGAGGACCTGTCTGAGGGACT 531
Db      120 -----ThrProSerSerThrAsnLysAspProProLeuThrGlnThr 133
Qy      532 CGGAGCCGG-----AGGGGGCCACCCCGGAGCGGACAGAGATCGGCACATTA 582
Db      134 ValGlnLysGluIleAspLysGlyValAsnAsnGlu-----ProLysSerAlaAsnIle 151
Qy      583 CCGCTGGCGACGTACGGCCCTCCACACCGGGGGCCAAATTCGACGCCCTCCAGTAT--- 639
Db      152 ProArgLeuCysProLeuGlnAlaValArgGlyGlyGluPheGlyProAlaArgValPro 171
Qy      640 TGGCCCTTTCTCTTCGAGATCTCTATAATTGGAATACTAACCATCCCTTTCTCGGAG 699
Db      172 ValProPheSerLeuSerAspLeuLysGlnIleLysIleAspLeuGlyLysPheSerAsp 191
Qy      700 GATCCCAACCCCTCAGCGGCTTGGTGAGTCCCTTATGTTCTCTACCCAGCTACTTGG 759
Db      192 AsnProAspGlyTyrIleAspValLeuGlnGlyLeuGlyGlnSerPheAspLeuThrTrp 211
Qy      760 GATGATTGTCAACAGCTCTCGACACTCTTCACACCGGAGCGGAGGAGAGAAATCTTA 819
Db      212 ArgAspIleMetLeuLeuLeuAsnGlnThrLeuThrProAsnGluArgSerAlaAlaVal 231
Qy      820 TTAGAGGCTAGAAA-----AATGTTCTGGGGCGGACGGCGGACCCACG 864
Db      232 ThrAlaAlaArgGluPheGlyAspLeuTrpTyrLeuSerGlnAlaAsnAsnArgMetThr 251
Qy      865 CGGTTGCCAAATGAGATTGATCGGA-----TTTCCTTTAACTCGCCCGGTTGG 915
Db      252 ThrGluGluArgThrProThrGlyGlnGlnAlaValProSerValAspProHisTrp 271
Qy      916 GACTTACACACGGCTGAGGTAGGAGAGCTTGAATAATCTATCCCGAGGCTCTGCTGGCG 975
Db      272 AspThrGluSerGluHisGlyAspTrpCysHisLysHisLeuLeuThrCysValLeuGlu 291
Qy      976 GGTCTCCGGGGCGCTCAAGACGCCCACTAATTTGGCTAAGGTAAGGAAGTATGTCAG 1035
Db      292 GlyLeuArgLysThrArgLysLysPheMetAsnTyrSerMetSerThrIleThrGln 311
Qy      1036 GGACCGAATGAACCCCTCTGTTTCTTGAGAGGCTCTTGGAAAGCCTTCAGCGGTAC 1095
Db      312 GlyLysGluGluAsnLeuThrAlaPheLeuAspArgLeuArgGluAlaLeuArgLysHis 331
Qy      1096 ACCCTTTTGTATCCACTCGAGGGCCCAAAAGCCCTCAGTGGCTTTGGCTTTATAGGA 1155
Db      332 ThrSerLeuSerProAspSerIleGluGlyGlnIleLeuLysAspLysPheIleThr 351

```

```

Qy      1156 CAGTCAGCCTTGGATATTAGAAAGAGGCTTCAGAGACTGGAAGGTTACAGGAGCT 1212
Db      352 GlnSerAlaAlaAspIleArgLysAsnPhelysSerLeuProLysLeuAlaAlaAla 370

```

Search completed: February 14, 2006, 16:15:34  
Job time : 68.6237 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: February 14, 2006, 14:51:21 ; Search time 47.4857 Seconds  
(without alignments)  
3295.451 Million cell updates/sec

Title: US-10-723-552-3  
Perfect score: 14636  
Sequence: 1 GCGTGGTGACGACTGTGGG.....CTGTTTGATCAAAAAAAA 8132

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSSWEB.spool/US10723552/runat.14022006.125143.12937/app.query.fasta.1  
-DB=PIR -QPMW=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCI=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p  
-USER=US10723552 @CGN 1.1.107 @runat.14022006.125143.12937 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5237	35.8	1784	2	Tl0532 gag-pol polyprotei
2	4316.5	29.5	1165	1	GNLJGL HIV-1 retropepsin
3	3999	27.3	1196	1	GNMVG HIV-1 retropepsin
4	3922.5	27.3	1204	2	S35475 pol polyprotei
5	3980.5	27.2	1204	2	S70393 pol polyprotei
6	3973.5	27.1	1199	1	GNMVL HIV-1 retropepsin
7	3947	27.0	1186	1	GNMVR HIV-1 retropepsin
8	3940.5	26.9	1189	1	GNMVM HIV-1 retropepsin
9	3536.5	24.2	1046	1	GNMVCE HIV-1 retropepsin
10	2639.5	18.0	843	1	GNVWK pol polyprotei
11	1862.5	12.7	559	2	A46311 pol polyprotei
12	1725.5	11.8	476	2	S04842 pol polyprotei
13	1681	11.5	581	2	A42743 pol polyprotei
14	1677.5	11.5	520	1	FOLJGL gag polyprotei

15	1643	11.2	1751	2	T09394 gag-pro-pol polypr
16	1637.5	11.2	512	1	FOMVGS env polyprotein -
17	1505.5	10.3	667	1	VCLJGL env polyprotein pr
18	1477.5	10.1	537	1	FOMVM7 gag polyprotein -
19	1383.5	9.5	538	2	S35474 gag polyprotein -
20	1381.5	9.4	538	2	S70394 gag polyprotein -
21	1379.5	9.4	538	1	FOMVM gag polyprotein -
22	1376.5	9.4	538	1	FOMVIM gag polyprotein -
23	1371	9.4	537	1	FOMVGV gag polyprotein -
24	1364	9.3	537	1	FOMVRV gag polyprotein -
25	1359.5	9.3	540	1	FOMVHL gag polyprotein -
26	1357	9.3	537	1	FOMVMB gag polyprotein -
27	1351	9.2	676	2	S70395 env polyprotein -
28	1347.5	9.2	536	1	FOMVME gag polyprotein -
29	1347	9.2	676	2	T01381 env protein - muri
30	1345	9.2	676	1	VCMPV env polyprotein pr
31	1344	9.2	665	1	VCMVKA env polyprotein pr
32	1339	9.1	665	1	VCMVVR env polyprotein pr
33	1339	9.1	669	1	VCVWEK env polyprotein -
34	1335	9.1	669	2	A46511 envelope protein -
35	1335	9.1	689	1	B43491 env polyprotein -
36	1332.5	9.1	688	2	A43491 env polyprotein -
37	1326	9.1	529	1	FOMVUU gag polyprotein -
38	1320	9.0	665	1	VCVWEM env polyprotein -
39	1313	9.0	666	1	VCMVHL env polyprotein pr
40	1312	9.0	532	1	FOMVVB gag polyprotein -
41	1290	8.8	640	1	VCMVRV env polyprotein pr
42	1288	8.8	661	1	VCMVCB env polyprotein -
43	1276	8.7	642	1	VCMVFG env polyprotein -
44	1273	8.7	642	2	Tl0533 env polyprotein pr
45	1268.5	8.7	536	1	FOMVMD gag polyprotein -

ALIGNMENTS

RESULT 1  
Tl0532

gag-pol polyprotein precursor - feline leukemia virus (strain FeLV-FAIDS)  
N;Contains: core protein p10; core protein p12; core protein p15; core protein p27; end  
C;Species: feline leukemia virus  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C;Accession: Tl0532  
J;Donahue, P.R.; Hoover, E.A.; Beltz, G.A.; Riedel, N.; Hirsch, V.M.; Overbaugh, J.; Mu  
R; Virol. 62, 722-731, 1988  
A;Title: Strong sequence conservation among horizontally transmissible, minimally patho  
A;Reference number: Z17078; MUID:88119207; PMID:2828667  
A;Accession: Tl0532  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-1784 <DON>  
A;Cross-references: UNIPROT:Q85521; UNIPARC:UPI000010C1P5; EMBL:M18247; NID:G3233904; PI  
C;Superfamily: pol polyprotein  
C;Keywords: hydrolase; nucleotidyltransferase; polyprotein; reverse transcriptase  
F;75-576/Product: gag polyprotein #status predicted <GAG>  
F;75-201/Product: core protein p15 #status predicted <P15>  
F;202-271/Product: core protein p12 #status predicted <P12>  
F;272-519/Product: core protein p27 #status predicted <P27>  
F;520-576/Product: core protein p10 #status predicted <P10>  
F;577-1784/Product: pol polyprotein #status predicted <POL>  
F;577-701/Product: proteinase #status predicted <PTN>  
F;702-1368/Product: RNA-directed DNA polymerase (EC 2.7.7.49) #status predicted <REV>  
F;1369-1784/Product: endonuclease (EC 3.1.1.-) #status predicted <EDC>

Alignment Scores:  
Pred. No.: 2,74e-314 Length: 1784  
Score: 5237.00 Matches: 1028  
Percent Similarity: 70.3% Conservative: 232  
Best Local Similarity: 57.3% Mismatches: 416  
Query Match: 35.8% Indels: 117  
DB: 2 Gaps: 27

US-10-723-552-3 (1-8132) x Tl0532 (1-1784)





Db 722 tAspIleTrpLeuLysAsnPheProGlnAlaTrpAlaGluThrGlyMetGlyMetal 742  
 Qy 2639 AAGCAAGTTCCCCCAAGATTATTCAACTGAAGCCAGTGCACACAGCAGTGTCAAGTCAG 2698  
 Db 742 aHisCysGlnAlaProValLeuIleGlnLeuLysAlaThrAlaThrProIleSerIleAr 762  
 Qy 2699 ACAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTTCGGCCGCGATGTCGAAGATTAAAT 2758  
 Db 762 gGlnTyrProMetProHisGlnAlaTyrGlnGlyIleLysProHisIleArgArgMetLe 782  
 Qy 2759 CCAACAGCGGATCTCTGCTGCTCCCAATCTCCCTGGAATACTCCCTGCTACTACCGGTTAG 2818  
 Db 782 uAspGlnGlyIleLeuLysProCysGlnSerProIlePheProIleLeuLeuProVally 802  
 Qy 2819 AAAGCTGGAGCTAATCACTATCGACAGTACAGGACTTCGAGAGAGTCAATAAACCGGT 2878  
 Db 802 sLysProGlyThrLysAspTyrArgProValGlnAspLeuArgGluValAsnLysArgVa 822  
 Qy 2879 GCAGGATATACACCCACAGTCCGGAACCTTATTAACCTCTTGCTGCTCTCCACCCCA 2938  
 Db 822 lGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerThrLeuProProSe 842  
 Qy 2939 ACGGAGCTGTATACGATTATGGACTTAAAGGATGCTCTCTGCTGAGATTACACCC 2998  
 Db 842 rHisProTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisSe 862  
 Qy 2999 CACTAGCCAAACCACTTTTCCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCCGGCA 3058  
 Db 862 rGluSerGlnLeuLeuPheAlaPheGluTrpArgAspProGluIleGlyLeuSerGlyGI 882  
 Qy 3059 GCTCACCCTGACCGGACTGCCAGAGGTTCAAGAACTCCCGACACATCTTTGACGAAGC 3118  
 Db 882 nLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAspGluAl 902  
 Qy 3119 CCTACACAGAGACTGSCCAACTTCAGGATCCAAACACCCTCAGTGACCTCCTCCAGTA 3178  
 Db 902 aLeuHisSerAspLeuAlaAspPheArgValArgTyrProAlaLeuValLeuLeuGlnTy 922  
 Qy 3179 CGTGGATGACTGCTCTTCGCGGGAGCCACCAACAGGACTGTAGAGGCCACGAAGGC 3238  
 Db 922 rValAspAspLeuLeuLeuAlaAlaAlaThrArgThrGluCysLeuGluGlyThrLysAl 942  
 Qy 3239 ACTACTGCTGGAAATTGTCTGACCTAGGCTACAGAGCTCTGCTAAGAGCCCGAGATTG 3298  
 Db 942 aLeuLeuGluThrLeuGlyAsnLysGlyTyrArgAlaSerAlaLysLysAlaGlnIleCy 962  
 Qy 3299 CAGGAGAGAGTAACATACTTGGGGTACATTTGCGGAGCGGCAGCGATGGCTGACGGA 3358  
 Db 962 sLeuGlnGluValThrTyrLeuGlyTyrSerLeuLysAspGlyGlnArgTrpLeuThrLy 982  
 Qy 3359 GGCACGGAAGAAACTGTAGTCCAGATACCGGCCCCCAACACACAGCCAAACAAATGAGAGA 3418  
 Db 982 sAlaArgLysGluAlaIleLeuSerIleProValProLysAsnProArgGlnValArgGI 1002  
 Qy 3419 GTTTTGGGACAGCTGGATTGTCAGCTAGGCTACAGAGCTCTGCTAAGAGCCCGAGATTG 3478  
 Db 1002 uPheLeuGlyThrAlaGlyTyrCysArgLeuTrpIleProGlyPheAlaGluLeuAlaI 1022  
 Qy 3479 CCACCTCTACCCGCTAACCAAGAAAGGGGAATTCCTCGGCTCCTGAGCACCAGAA 3538  
 Db 1022 aProLeuTyrProLeuThrArgProGlyThrLeuPheGlnTrpGlyThrGlnGlnLe 1042  
 Qy 3539 GGCATTGTGATGCTATCAAAAGGCCCTGTGAGCGCACCTGCTCTCGCCCTCCCTGACCT 3598  
 Db 1042 uAlaPheGluAsnIleArgLysAlaLeuLeuSerProAlaLeuGlyLeuProAspIle 1062  
 Qy 3599 AACTAAACCTTTACCTTTATGTGATGAGCTAAGGAGTACCCCGGGAGTTTAAAC 3658  
 Db 1062 eThrLysProPheGluLeuPheIleAspGluAsnSerGlyPheAlaLysGlyValLeuVa 1082  
 Qy 3659 CCNAACCTTAGGACCTGAGAGACCTGTGCTGCTACCTGTCAAGAACCTGATCCTGT 3718  
 Db 1082 lGlnLysLeuGlyProTrpLysArgProValAlaTyrLeuSerLysLysLeuAspThrVa 1102

Qy 3719 AGCAGTGGTGGCCCATATGCTGAAGGCTATCGCAGCTATGCGCCATCTGCTGCTCAAGGA 3778  
 Db 1102 lAlaSerGlyTrpProCysLeuArgMetValAlaAlaIleAlaIleLeuValLysAs 1122  
 Qy 3779 CGCTGCAAAATTTGACTTTGGGACAGAAATATACTGTAATAGCCCCCATGATTTGGAGAA 3838  
 Db 1122 pAlaGlyLysLeuThrLeuGlyGlnProLeuThrIleLeuThrSerHisProValGluAl 1142  
 Qy 3839 CATCGTTCCGACGCCCCAGACCGGATGAGCAACGCCCGCATGACCCACCTATCAAG 3898  
 Db 1142 aLeuValArgGlnProProAsnLysIlePheSerAsnAlaArgMetThrHisIleGlnAl 1162  
 Qy 3899 CTGCTTCTTC- --ACAGAGAGGGTCACTGTTCCGCTCCACCGCCGCTCTCAACCTCTGCCAC 3955  
 Db 1162 aMetLeuLeuAspAlaGluArgValHisPheGlyProThrValSerLeuAsnProAlaTh 1182  
 Qy 3956 TCTTCTGCTGGAAGACTGATGAACCA- ----GTGACTCATGATTGCCACTCACTATT 4009  
 Db 1182 rLeuLeuProLeuProSerGlyLysProProArgLeuSerProAsp- ---- 1197  
 Qy 4010 GATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTCGGAGAGTCT 4069  
 Db 1198 -LeuAlaGluThrMetAlaGlnThrAspLeuThrAspGlnProLeuProAspAlaAspLe 1217  
 Qy 4070 AACCTGTTTCACTGACGGAAGCAGCTATGTGTGGAGGTAAGAGGATGCTCGGGCGGC 4129  
 Db 1217 uThrTrpTyrThrAspGlySerSerPheIleArgAsnGlyGluArgLysAlaGlyAlaAl 1237  
 Qy 4130 GGTGGTGGACGGGACCCGACGATCTGGGCCAGCAGCTGCGGAGAGAACTTTCAGACA 4189  
 Db 1237 aValThrThrGluSerGluValIleTrpAlaAlaSerLeuProProGlyThrSerAlaGI 1257  
 Qy 4190 AAAGGCTGAGCTCATGCGCCTCAGCAAGCTTTGCGCTGCGCGAGGGAATCCATAA 4249  
 Db 1257 nArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaLysGlyLysLeuTh 1277  
 Qy 4250 CATTTATACGGACAGCAGGTATGCTTTCGACTGCACACGCTACATCGGGGCCATCTATAA 4309  
 Db 1277 rValTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyGluIleTyrAr 1297  
 Qy 4310 ACNAAGGGGTGCTTACCTCCTCAGCAGGAGGAGGAATAAAGACAAAGAGAAATTTCTAAG 4369  
 Db 1297 gArgArgGlyLeuLeuThrSerGluGlyLysGluIleLysAsnLysAsnGluIleLeuAl 1317  
 Qy 4370 CTTATTAGAGCGGTACATTTTACCAAAAGGCTAGCTATTATACCTGCTCTCGACATCA 4429  
 Db 1317 aLeuLeuGluAlaLeuPheLeuProLysArgLeuSerIleIleHisCysProGlyHisGI 1337  
 Qy 4430 GAAAGCTAAAGATCTCATATCCAGAGGAACACAGATGGCTGACCGGTTGCCAAGCAGGC 4489  
 Db 1337 nLysGlyAspSerProGlnAlaLysGlyAsnArgLeuAlaAspAspThrAlaLysLysAl 1357  
 Qy 4490 AGCC- ----CAGGGTGTAACTCTTCGCT- ----ATAATAGAAATGCCCA 4531  
 Db 1357 aAlaThrGluThrGlnSerSerLeuThrIleLeuProThrGluLeuIleGluGlyProLy 1377  
 Qy 4532 AGCCCCAGAACCCAGACGACGACTACACCTTAGAGACTGCGAAGAGATATAAAGATAGA 4591  
 Db 1377 s- ----ArgProProTrpGluTyrAspAspSerAspLeuAspLeuValGlnLysLeuGI 1395  
 Qy 4592 CCAGTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATCTCGCC 4651  
 Db 1395 uAlaHisTyrGluProLysArgGlyThr- --TrpGluTyrArgGlyLysThrIleMetPr 1414  
 Qy 4652 CCACAAAGAGGGTTAGAAATATGTCCACAGATACATGCTTAACCCACCTAGGAATAA 4711  
 Db 1414 oGluLysTyrAlaLysGluLeuIleSerHisLeuHisLysLeuThrHisLeuSerAlaAr 1434  
 Qy 4712 ACACCTGACGAGTGGTTCAGAACATCC- ----CCTTATCATGTTCT 4753  
 Db 1434 gLysMetLysThrLeuLeuGluArgGluThrGlyPheTyrLeuProAsnArgAspLe 1454



Db 220 LysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluIleAsnLysArgVal 239  
QY 2880 CAGGATATACACCAACAGCCCAAGACCTTATAACCTCTGTGTGCTCTCCACCCCAA 2939  
Db 240 GlnAspIleHisProThrValProAsnProTyrAsnLeuLeuSerSerLeuProProSer 259  
QY 2940 CGGAGCTGGTATACAGTATTGGACTTAAAGATGCCCTTCTCTGCTGAGATTACACCCC 2999  
Db 260 TyrThrTyrSerValLeuAspLeuLysAspAlaPheCysLeuArgLeuHisPro 279  
QY 3000 ACTAGCCAAACCTTTTGGCTTCGAATGAGAGATCCAGTACCGGAAGAACCGGGCAG 3059  
Db 280 AsnSerGlnProLeuPheAlaPheGluTrpLysAspProGluLysGlyAsnThrGlyGln 299  
QY 3060 CTACCTGGACCCGACTGCCCAAGGGTTCAGAACTCCCGACCATCTTTGACGAAGCC 3119  
Db 300 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAspGluAla 319  
QY 3120 CTACACAGACCTGGCCAACTTCAGGATCCAAACACCTCAGGTGACCTCTCCAGTAC 3179  
Db 320 LeuHisArgAspLeuAlaProPheArgAlaLeuAsnProGlnValValLeuLeuGlnTyr 339  
QY 3180 GTGGATGACCTGTTCTGGGGGAGGCCACCAACAGAGACTGCTTAGAGCACAAGGCCA 3239  
Db 340 ValAspAspLeuLeuValAlaAlaProThrTyrGluAspCysLysLysGlyThrGlnLys 359  
QY 3240 CTACTGCTGAAATTGCTGACCTAGGCTACAGAGCTCTGCTAAGAGGCCACAGATTGC 3299  
Db 360 LeuLeuGlnLeuSerLysLeuGlyTyrArgValSerAlaLysLysAlaGlnLeuCys 379  
QY 3300 AGGAGAGGTAAACATCTTTGGGGTACAGTTTGGGGACGGCAGCGATGGCTGACGGAG 3359  
Db 380 GlnArgGluValThrTyrLeuGlyTyrLeuLeuLysGluGlyLysArgTrpLeuThrPro 399  
QY 3360 GCACGGAGAACTGTAGTCCAGATACCGGCCCAACACACAGCCAAACAAATGACAGAG 3419  
Db 400 AlaArgLysAlaThrValMetLysIleProValProThrThrProArgGlnValArgGlu 419  
QY 3420 TTTTGGGGACAGCTGGATTTCGAGCTGTGATCCCGGGTTTGGACTTAGCAGCC 3479  
Db 420 PheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaSerLeuAlaAla 439  
QY 3480 CCACCTACCCGCTAACCAAGAAAAGGGAAATTCCTCGGGCTCTCGAGCACCAAG 3539  
Db 440 ProLeuTyrProLeuThrLysGluSerIleProPheIleTrpThrGluGluHisGlnGln 459  
QY 3540 GCATTTGATGCTATCAAAAGGCCCTGCTGAGGCGACCTGCTCTGGCCCTCCCTGACGTA 3599  
Db 460 AlaPheAspHisIleLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspLeu 479  
QY 3600 ACTAAACCCCTTTACCTTTATGTGGATGACGCTAAGGGAGTAGCCCGGGAGTTTAAAC 3659  
Db 480 ThrLysProPheThrLeuTyrIleAspGluArgAlaGlyValAlaArgGlyValLeuThr 499  
QY 3660 CAAACCTTAGGACCATGGAGAAGACCTGTGCTCTACCTGTCAAGAAAGCTCGATCTGTA 3719  
Db 500 GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLeuAspProVal 519  
QY 3720 GCCAGTGGTGGCCCATATGCCTAAGGCTATCGAGCTGTGGCCATATGTTGCAAGGAC 3779  
Db 520 AlaSerGlyTrpProThrCysLeuLysAlaValAlaAlaValAlaLeuLeuLysAsp 539  
QY 3780 GCTGACAAATTGACTTTGGACAGAATATACTGTATAGCCCTCCATGCTGGAGAAC 3839  
Db 540 AlaAspLysLeuThrLeuGlyGlnAsnValThrValIleAlaSerHisSerLeuGluSer\*559  
QY 3840 ATCGTTGGCAGCCCGACGATCGATGACCAACCGCCGATGACCCACCTATCAAGC 3899  
Db 560 IleValArgGlnProProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSer 579  
QY 3900 CTGCTTCTCACAGAGGGTACGTTGCTTCCACCGCCGCTCTCAACCTCGCACTCTT 3959

Db 580 LeuLeuLeuAsnGluArgValSerPheAlaProProAlaValLeuAsnProAlaThrLeu 599  
QY 3960 CTCCTCGAAGAGACTGATGAACACAGTCACTGATGTCCTCACTCACTATTGATTGAGGAG 4019  
Db 600 LeuProValGluSerGluAlaThrProValHisArgCysSerGluIleLeuAlaGluGlu 619  
QY 4020 ACTGGGGTCCGAAGGACCTTACAGACATACCCCTGACTGCTGAGAAAGTGTAACTGGTTC 4079  
Db 620 ThrGlyThrArgArgAspLeuGluAspGlnProLeuProGly--ValProThrTyr 638  
QY 4080 ACTGACCGAAGACGCTATGTGGTGAAGGTAAAGAGTGGCTGGGGCGGGTGGTGCAGC 4139  
Db 639 ThrAspGlySerSerPheIleThrGluGlyLysArgAlaGlyAlaProIleValAsp 658  
QY 4140 GGCACCGCAGCATCTGGCCAGCAGCCTGCCGGAAGGAATCTTCAGCACAAAAGCTGAG 4199  
Db 659 GlyLysArgThrValTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 678  
QY 4200 CTCATGCCCTCACGCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAACATTTATACG 4259  
Db 679 LeuValAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysAsnIleAsnIleTyrThr 698  
QY 4260 GACAGCAGGTATGCCCTTTCGACTGCACAGTACATGCGGCCATCTTATAACAAAGGGGG 4319  
Db 699 AspSerArgTyrAlaPheAlaThrAlaHisIleHisGlyAlaIleTyrLysGlnArgGly 718  
QY 4320 TTGCTTACCTCAGCAGGAGGAAATAAAGAACAAAGAGGAATTCCTAAGCCTATTAGAA 4379  
Db 719 LeuLeuThrSerAlaGlyLysAspIleLysAsnLysGluGluIleLeuAlaLeuLeuGlu 738  
QY 4380 GCCGTACATTACCAAAAAGCTTATATACACTGTCTCGACATCAGAAAGCTATAA 4439  
Db 739 AlaIleHisLeuProArgValAlaIleIleHisCysProGlyHisGlnArgGlySer 758  
QY 4440 GATCTCATATCCAGAGAAACCGATGGCTGACCGGGTTCGCAAGCAGGAGCCAGCGT 4499  
Db 759 AsnProValAlaThrGlyAsnArgAlaAspGluAlaAlaLysGlnAlaLeuSer 778  
QY 4500 GTTAACTTCTGCTATAATAAGAAATGCCAAAGCCCCA--CAACCCAGCACACAGTAC 4556  
Db 779 ThrArgValLeuAlaGlyThrThrLysProGlnGluProIleGluProAlaGlnGluLys 798  
QY 4557 ACCTTAGAAGACTGGCAAGAGATAAANAAGATAGACCAGTTCTCTGAGACTCCGGAAGG 4616  
Db 799 ThrArgPro----- 801  
QY 4617 ACCTGCTATACCTCAGATGGGAAGAAATCTGCTGCCCAACAAAGAGGTTAGAAATGTC 4676  
Db 802 -----ArgGluLeuThrProAspArg--GlyLysGluPheIle 813  
QY 4677 CAACAGATACATCGTAAACCCACCTAGGAACCTAAACACCTGCAGCAGTTGGTCAAGAACA 4736  
Db 814 LysArgLeuHisGlnLeuThrHisLeuGlyProGluLysLeuLeuGlnLeuValAsnArg 833  
QY 4737 TCCCTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTCGGTGCTCAACATTTGTGTG 4796  
Db 834 ThrSerLeuLeuIleProAsnLeuGlnSerAlaValArgGluValThrSerGlnCysGln 853  
QY 4797 CCTGTCAGCTGGTTAATCTAATCTCTCCAGATCCCTCCAGGGAAGAGACATAAGGGA 4856  
Db 854 AlaCysAlaMetThrAsnAlaValThrThrTyrArgGluThrGlyLysArgGlnArgGly 873  
QY 4857 AGCCACCCAGGCGCTCACCTGGGAAGTGGACTTCCTCAGTGAAGCGGCTAATAACGGA 4916  
Db 874 AspArgProGlyValTyrTrpGluValAspPheThrGluIleLysProGlyArgTyrGly 893  
QY 4917 AACAAATACCTATTGTTTGTAGACACCTTTTTCAGGATGGGTAGAGGCTTATCTACT 4976  
Db 894 AsnLysTyrLeuLeuValPheIleAspThrPheSerGlyTrpValGluAlaPheProThr 913  
QY 4977 AAGAAGAGACTTCAACCTGGTGGCTAAATAAATACTCGAAGAAATTTTCAAGATTT 5036  
Db 914 LysThrGluThrAlaLeuIleValCysLysLysIleLeuGluGluIleLeuProArgPhe 933

[illegible]

C;Comment: This protein is synthesized as a gag-pol polyprotein.  
C;Comment: The pol polyprotein contains reverse transcriptase and possibly a nuclease or RNase.  
C;Genetics:   
A;Gene: pol  
C;Keywords: aspartic proteinase; hydrolase; nucleotidyltransferase; polyprotein; reverse transcriptase  
F:3-102/Product: retropepsin #status predicted <RTP>  
F:27/Active site: Asp (shared with dimeric partner) #status predicted

QY 2994 CACCCCTAGCCCAACCACTTTTGGCTTCGAATGGAGAGATCCAGGTACGGGAAGAACC 3053  
DB HisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyMetGlyIleSer 300  
QY 3054 GGGCAGCTCACCTCGACCGACTGCCCCAGGGTTCAAGACTCCCGACCATCTTTGAC 3113  
DB GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheIysAsnSerProThrLeuPheAsp 320  
QY 3114 GAAGCCCTACACAGAGACTGGCCCAACTTCAGGATCCAAACACCTCAGGTGACCCCTCCTC 3173  
DB GluAlaLeuHisArgAspLeuAlaAspPheArgIleGlnHisProAspLeuIleLeuLeu 340  
QY 3174 CAGTACCTGATGACCTGCTTCTGGCGGAGCCACCAACAGGACTGCTTAGAAGGCACG 3233  
DB GlnTyrValAspAspIleLeuLeuAlaAlaThrSerGluLeuAspCysGlnGlnGlyThr 360  
QY 3234 AAGGCACCTACTGCTGGAATGCTCGACCTAGGCTACAGAGCCTCTGCTAGNAGGCCCGCAG 3293  
DB ArgAlaLeuLeuLeuThrLeuGlyAsnLeuGlyTyrArgAlaSerAlaIleLysAlaGln 380  
QY 3294 ATTTGCAGGAGAGAGGTAAACATACTTCGGGTACAGTTTGGGGACGGGACGCGATGGCTG 3353  
DB LeuCysGlnLysGlnValLysTyrLeuGlyTyrLeuLeuLysGluGlyGlnArgTrpLeu 400  
QY 3354 ACGGAGCGACGGAAGAAACTGTAGTCCAGATACCGGCCCCCAACACCAAGCAACAAATG 3413  
DB ThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGlnLeu 420  
QY 3414 AGAGATTTTGGGACAGCTGGATTTTGGAGACTGTGGATCCCGGGTTTGGACCTTA 3473  
DB ArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGluMet 440  
QY 3474 GCAGCCCACTTACCCGCTAAACCAAGAAAAGGGGAATTCCTGGGCTCCTGAGCAC 3533  
DB AlaAlaProLeuTyrProLeuThrLysThrGlyThrLeuPheAsnTrpGlyProAspGln 460  
QY 3534 CAGAAGCATTTGATGCTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCT 3593  
DB GlnLysAlaTyrGlnGluIleLysGlnAlaLeuLeuThrAlaProAlaLeuGlyLeuPro 480  
QY 3594 GAGTAACTAAACCTTTACCTTTATGTGGATGAGGTAAGGAGTAGCCCGGAGATT 3653  
DB AspLeuThrLysProPheGluLeuPheValAspGluLysGlnGlyTyrAlaLysGlyVal 500  
QY 3654 TTAACCCAAACCTTAGGACCATGGAGAAGCTGTGCGCTACCTGTCAAGAAGAGCTCGAT 3713  
DB LeuThrGlnLysLeuGlyProThrArgArgProValAlaTyrLeuSerLysLysLeuAsp 520  
QY 3714 CTTGTAGCCAGTGGTGGCCCATATGCTGGAAGGCTATCGCAGCTGTGGCCATCTGGTC 3773  
DB ProValAlaAlaGlyTrpProProCysLeuArgMetValAlaAlaIleAlaValLeuThr 540  
QY 3774 AAGGAGCTGACAAATGACTTTGGGACAGAAATATTAAGTAAATAGCCCGCCCATGATTG 3833  
DB LysAspAlaGlyLysLeuThrMetGlyGlnProLeuValIleLeuAlaProHisAlaVal 560  
QY 3834 GAGAACATCTCTCGGAGCCCGCCAGACCGATGATGACCAACCGCCGATGACCCACTAT 3893  
DB GluAlaLeuValLysGlnProProAspArgTrpLeuSerAsnAlaArgMetThrHisTyr 580  
QY 3894 CAAAGCCTGCTCTC---ACAGAGAGGGTCCAGCTTCGCTCCACAGCGCTCTCAACCCCT 3950  
DB GlnAlaMetLeuLeuAspThrAspArgValGlnPheGlyProValValAlaLeuAsnPro 600  
QY 3951 GCCACTCTTCTGCTGAAGAGACTGATGAACACAGTGAATGATGATGCCATCACTATTG 4010  
DB AlaThrLeuLeuPro---LeuProGluGluGlyAlaProHisAspCysLeuGluIleLeu 619  
QY 4011 ATTGAGGAGACTGGGCTCGCAGGACCTTACAGACATACCTGCTGAGAGAGTGCTA 4070  
DB AlaGluThrHisGlyThrArgProAspLeuThrAspGlnProIleProAspAlaAspHis 639

QY 4071 ACCTGGTTCTACTGACGGAACGACGCTATGTGTGGAAAGTAAGAGATGGCTGGCGCGCG 4130  
DB ThrTrpTyrThrAspGlySerSerPheLeuGlnGlnGlyGlnArgLysAlaGlyAlaAla 659  
QY 4131 GTGGTGAACGGGACCGCCACGATCTGGCCAGACGCTGCCGGAAGAACTTCAGACAA 4190  
DB ValThrThrGluThrGluValIleTrpAlaArgAlaLeuProAlaGlyThrSerAlaGln 679  
QY 4191 AAGCCTGAGCTCATGGCCCTCACCAAGCTTTGGGCTGGCCGAAGGAAATCCATAAC 4250  
DB ArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLysArgLeuAsn 699  
QY 4251 ATTTATACGACGACAGGTATGCTTTGCGACTGCACACGTACATGCGGCCATCTATAA 4310  
DB ValTyrThrAspSerArgTyrAlaPheAlaThrAlaHisIleHisGlyGluIleTyrArg 719  
QY 4311 CAAAGGGGTGCTTACCTCAGCAGGAGGGAATTAAGAAACAAGAGGAATTTCAAGC 4370  
DB ArgArgGlyLeuLeuThrSerGluGlyArgGluIleLysAsnLysSerGluIleLeuAla 739  
QY 4371 CTATTAGAACCCGTACATTTACCAAAAGCTAGCTATTATACACTCTCTGGACATCAG 4430  
DB LeuLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysLeuGlyHisGln 759  
QY 4431 AAAGCTAAAGATCTATATCCAGAGGAACACAGATGGCTGACCGGTTGCCAAGCAGCA 4490  
DB LysGlyAspSerAlaGluAlaArgGlyAsnArgLeuAlaAspGlnAlaAlaArgGluAla 779  
QY 4491 GCC-----CAGGGTGTAACTTCTGCTTATATAGAAATGCCCAAGCCCA 4538  
DB AlaIleLysThrProProAspThrSerThrLeuLeuIleGluAspSerThrProTyrThr 799  
QY 4539 GAACCCAGACGACAGTACACCTTAGAAGCTGCCAGAGATATAAAGATAGACCACTTC 4598  
DB ProAlaTyrPheHisTyrThrGluThrAspLeuLysLysLeuArgGluLeuGlyAlaThr 819  
QY 4599 TCTGAGACTCCGGAAGGACCTCTATACCTCAGATGGGAAGAAATCTCTGCCCAAAA 4658  
DB TyrAsnGlnSerLysGly--TyrTrpValPheGlnGlyLysProValMetProAspGln 838  
QY 4659 GAAGGTTAGATATATCCACAGATACATCTGTCTAACCCACCTAGGAACCTAACACCTG 4718  
DB PheValPheGluLeuLeuAspSerLeuHisArgLeuThrHisLeuGlyTyrGlnLysMet 858  
QY 4719 CAGCAGTGTGTC-----AGAACATCCCTTATCATGTTCTGAGGCTTACCAGGAGTGGCT 4772  
DB LysAlaLeuLeuAspArgGlyGluSerProTyrTyrMetLeuAsnArgAspLysThrLeu 878  
QY 4773 GACTCGGTGTCAAAACATTTGTGTCCTGCTGACCTGTTAACTTAATCTCTCCAGAAATG 4832  
DB GlnTyrValAlaAspSerCysThrValCysAlaGlnValAsnAlaSerLysAlaLysIle 898  
QY 4833 CCTCCAGGAAGAGACTAAGGGGAAGCCACCCAGGCGCTCAGTGGGAAGTGGACTTCACT 4892  
DB GlyAlaGlyValArgValArgGlyHisArgProGlySerHisTrpGluIleAspPheThr 918  
QY 4893 GAGGTAAAGCCGCTAAATACCGAAACAATACCTATTGTTTCTGAGACACTTTCA 4952  
DB GluValLysProGlyLeuTyrGlyTyrLysTyrLeuValPheValAspThrPheSer 938  
QY 4953 GGATGGGTAGAGCTTATCTACTAAGAAAGAGACTTCAACCGTGGCTGGCTAAAAAATA 5012  
DB GlyTrpValGluAlaPheProThrLysArgGluThrAlaArgValValSerLysLysLeu 958  
QY 5013 CTGGAAGAAATTTTCCAAAGATTTGGAATACCTAAGTAAATAGGGTCAGACAATGGTCCA 5072  
DB LeuGluGluIlePheProArgPheGlyMetProGlnValLeuGlySerAspAsnGlyPro 978  
QY 5073 GCTTTTGTGTCAGTAACTCAGGACTGGCCAGATATTGGGATTTGATTTGGAACCTG 5132  
DB AlaPheThrSerGlnValSerGlnSerValAlaAspLeuLeuGlyIleAspTrpLysLeu 998  
QY 5133 CATTGTGCATACAGACCCCAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATTA 5192

```

Db      999  HisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLys 1018
Qy      5193  GAGACCTTACTAAATGACCGCGAGAGCTGGCGTTAATGATGGATAGCTCTCTGCC 5252
Db      1019  GluThrLeuThrLysLeuThrLeuAlaAlaGlyThrArgAspTrpValLeuLeuPro 1038
Qy      5253  TTTGTGCTTTTAGGGTTAGGAACACCCCTGGACAGATTTGGCTGACCCCTATGAATTA 5312
Db      1039  LeuAlaLeuTyrArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGluile 1058
Qy      5313  CTCTACGGGGGACCCCGCCCTAGTAAATTTCTTGTACATAGTGTGCTGAGCTGTG 5372
Db      1059  LeuTyrGlyAlaProProLeuValAsnPheHisAspProAspMetSerGluLeuThr 1078
Qy      5373  CTTTCCAGCCTTTGTTCTCTAGGCTCAAGGCACTTGAAGTGGTGGAGACAGCGCTGG 5432
Db      1079  AsnSerProSerLeuGlnAlaHisLeuGlnAlaLeuGlnThrValGlnArgGluIleTrp 1098
Qy      5433  AGGCAACTCCGGGAGGCTACTCAGGAGGAGGAGACTTGCAG---ATCCACATCGTTTC 5489
Db      1099  LysProLeuAlaGluAlaTyrArgAspGlnLeuAspGlnProValIleProHisProPhe 1118
Qy      5490  CAAGTGGGAGATTGAGTCTAGCTTAGAGCGCCACCGTGCAGGAAACCTCGAGACTCGGTGG 5549
Db      1119  ArgIleGlyAspSerValTrpValArgHisGlnThrLysAsnLeuGluProArgTrp 1138
Qy      5550  AAGGCGCCTTATCTGCTATCTTTTCCACACCAACGCTGTGAAAGTCGAAGGAATCTCC 5609
Db      1139  LysGlyProTyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleSer 1158
Qy      5610  ACCTGGATCCATGCTCCAGTTAAACCGCG-----CCACCTCCCGATTCGGGGTGG 5663
Db      1159  AlaTrpIleHisAlaAlaHisValLysAlaAlaThrThrProProIleLysProSerTrp 1178
Qy      5664  AAAGCCCAAGAGACTGAATCCCTTAAGCTTCGCTCCATCGCGTTCCT 5717
Db      1179  ArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGlyAlaPro 1196

RESULT 4
S35475
pol polyprotein - Friend murine leukemia virus (strain PVC-211)
N;Contains: nuclease; proteinase; reverse transcriptase
C;Species: Friend murine leukemia virus
A;Variety: strain PVC-211
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S35475
R;Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.; Masuda, M.
Nucleic Acids Res. 20, 3249, 1992
A;Title: Complete nucleotide sequence of a neuropathogenic variant of Friend murine leuk
A;Reference number: S35474; MUID:92319660; PMID:1620621
A;Accession: S35475
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: genomic RNA
A;Residues: 1-1204 <REM>
A;Cross-references: UNIPROT:P26808; UNIPARC:UPI0000131F06; EMBL:M93134; NID:G331898; PII
A;Experimental source: strain PVC-211
A;Note: the authors translated the stop codon at position 5 as Gln
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: polyprotein; reverse transcriptase

Alignment Scores:
Pred. No.: 1,03e-237 Length: 1204
Score: 3992.50 Matches: 757
Percent Similarity: 76.1% Conservative: 162
Best Local Similarity: 62.7% Mismatches: 257
Query Match: 27.3% Indels: 31
DB: 2 Gaps: 11

US-10-723-552-3 (1-8132) x S35475 (1-1204)

```

```

Qy      2154  GATTAGGGGAGACGGGTTTCGGACCCCTCCCGAGCCAGCGTAACCTTTGAAGTGGAG 2213
Db      4      AspGlnGlyGlyGlnGlyGlnProProGluProGluProGluProGluThrLeuLysValGly 23
Qy      2214  GGGCAACCACTGAGTTCCCTGGTTGATACCGGAGCGAAACATTCAGTGTCTACACCCA 2273
Db      24      GlyGlnProValThrPheLeuValAspThrGlyAlaGlnHisSerValLeuThrGlnAsn 43
Qy      2274  TTAGGAAACTAAAGATAAAATCCTGGGTGATGGTGGCCACAGGCGCAACACAGTAT 2333
Db      44      ProGlyProLeuSerAspLysSerAlaTrpValGlnGlyAlaThrGlyGlyLysArgTyr 63
Qy      2334  CCATGGACTACCCGAAACACAGTTGACTTCGGAGTGGGACGGGTAAACCACTCGTTCTG 2393
Db      64      ArgTrpThrThrAspArgValHisLeuAlaThrGlyLysValThrHisSerPheLeu 83
Qy      2394  GTCATACCTGAGTCCCGACACCCCTCTTATAGGTAGAGACTTATTGACCAAGATGGGAGCA 2453
Db      84      HisValProAspCysProTyrProLeuLeuGlyArgAspLeuLeuThrLysLeuAla 103
Qy      2454  CAAATTTCTTTTGA---CAAGGAAACACAGAGTGTCTCCTCAATAATAACAACCTATCACT 2510
Db      104      GlnIleHisPheGluGlySerGlyAlaGlnValValGlyProMetGlyGlnProLeuGln 123
Qy      2511  GTCTTGACCTCCCAATTAGATGACGAATATCGACTATACTCTCCCTAGTAAAGCCCTGAT 2570
Db      124      ValLeuThrLeuAsnIleGluAspGluTyrArgLeuHisGluThrSerLysGlyProAsp 143
Qy      2571  CAAAATATA---CAATTTCTGTTGGAAACAGTTTCCCCAACGCTGGGAGAGAAACCGCAGGG 2627
Db      144      ValProLeuGlySerThrTrpLeuSerAspPheProGlnAlaTrpAlaGluThrGlyGly 163
Qy      2628  ATGGTTTGGCAAGCAAGTTTCCCCCAAGTTTATTCAACTGAAGGCCAGTCCACACCA 2687
Db      164      MetGlyLeuAlaValArgGlnAlaProLeuIleLeuProLeuArgAlaIleSerThrPro 183
Qy      2688  GTGTCTAGTCAGACAGTACCCCTTGAGTAAAGAGAGTCAAGAGAAATTCGCCCGCATGTC 2747
Db      184      ValSerIleLysGlnTyrProMetSerArgGluAlaArgLeuGlyIleLysProHisIle 203
Qy      2748  CAAAGATTAATCAACAGGCGATCTCTAGTTCTCTGTCCAATCTCCCTGGAATATCCCGTG 2807
Db      204      GlnArgLeuLeuAspGlnGlyIleLeuValProCysGlnSerProTrpAsnThrProLeu 223
Qy      2808  CTACCGTTAGAAAGCTGGGACTAATGACTATCGACCACTACAGGCTTACAGGAGAGTGC 2867
Db      224      LeuProValLysLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluVal 243
Qy      2868  AATAAAGCGGTGACGAGTATACACCCCAAGTCCCGAACCTTATAACCTCTTGTGTGCT 2927
Db      244      AsnLysArgValGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerGly 263
Qy      2928  CTCCACCCCAACGGAGCTGTATACAGTATGAGATTAAGGATGCTCTTCTCTGCTGCTG 2987
Db      264      LeuProProSerHisGlnTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeu 283
Qy      2988  AGATTACACCCCACTAGCCCAACCACTTTTGGCTTCCCAATGGAGAGATCCAGGTACGGGA 3047
Db      284      ArgLeuHisProThrSerGlnSerLeuPheAlaPheGluTrpArgAspProGluMetGly 303
Qy      3048  AGAACCGGCGAGCTCACTCGACCCGACTCCCAAGGTTCAAGAACTCCCGACCATC 3107
Db      304      IleSerGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeu 323
Qy      3108  TTTGACGAAGCCCTTACACAGAGACCTGGCCCACTTCAGGATCCAAACACCTTCAGGTGACC 3167
Db      324      PheAspGluAlaLeuHisArgAspLeuAlaAspPheArgIleGlnHisProAspLeuIle 343
Qy      3168  CTCCTCAGTACGTGGATGACCTGCTTCTGCGGGAGGCCACCAACAGGACTGCTTAGAA 3227
Db      344      LeuLeuGlnTyrValAspAspLeuLeuAlaAlaThrSerGluLeuAspCysGlnGln 363

```









[illegible]

```
QY 5601 GGAATCTCCACCTGGATCCATCCACGTTAAACCGGCG-----CCA 5645
Db 1161 GlytlealaalaTPIleHisAlaalaHisVallysAlaalaAspThrArgIleGluPro 1180
QY 5646 CCTCCCGATTCGGGTGGAAAGCCGAAAGACTGAAATCCCTTAAGCTTCGCTCCAT 5705
Db 1181 ProSerGluSerThrTTPArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThr 1200
QY 5706 CGC 5708
Db 1201 Arg 1201

RESULT 6
GNMWIM
HIV-1 retropepsin (EC 3.4.23.16) - Moloney murine leukemia virus
N;Contains: nuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA polymera
C;Species: Moloney murine leukemia virus
A;Note: host Mus spp. (mouse)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 03-Jun-2002
C;Accession: A03956
R;Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.
Nature 293, 543-548, 1981
A;Title: Nucleotide sequence of Moloney murine leukaemia virus.
A;Reference number: A93265; MUID:82035843; PMID:6169994
A;Accession: A03956
A;Molecule type: genomic RNA
A;Residues: 1-1199 <SHI>
A;Cross-references: UNIPARC:UPI0000174A30
A;Experimental source: clone pMLV-1
A;Note: the pol polyprotein contains reverse transcriptase (about 80,000 daltons) and po
t yet been defined
C;Comment: This protein is synthesized as a gag-pol polyprotein.
C;Genetics:
A;Gene: pol
C;Superfamily: pol polyprotein
C;Keywords: aspartic proteinase; hydrolase; nucleotidyltransferase; polyprotein; reverse
F;3-102/Product: retropepsin #status predicted <RTP>
F;27/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:
Pred. No.: 1.52e-236 Length: 1199
Score: 3973.50 Matches: 753
Percent Similarity: 76.3% Conservative: 169
Best Local Similarity: 62.3% Mismatches: 254
Query Match: 27.1% Indels: 33
DB: 1 Gaps: 12

US-10-723-552-3 (1-8132) x GNMVIM (1-1199)
QY 2160 GGGAGACGGGTTCCGGACCCCTCCCGAGCCCGAGGTAACTTTGAAGTGGGGGCAA 2219
Db 1 GlyGlyGlnGlyGlnAspProProGluProArgIleThrLeuLysValGlyGln 20
QY 2220 CCAGTTCAGTTCCTGGTTGATACCGGACGCAAAATTCAGTGTACTACAGCCATTAGGA 2279
Db 21 ProValThrPheLeuValAspThrGlyAlaGlnHisSerValLeuThrGlnAsnProGly 40
QY 2280 AAACATAAAGATAAAAAATCCTGGGTGATGGGTGCCACAGGCGCAACACAGTATCCATGG 2339
Db 41 ProLeuSerAspLysSerAlaTIPValGlnGlyAlaThrGlyGlyLysArgTyrArgTip 60
QY 2340 ACTACCGAAGAACAGTTGACTTGGGAGTGGGAGTGAACCCACTGTTTCTGGTCATA 2399
Db 61 ThrThrAspArgLysValHisLeuAlaThrGlyLysValThrHisSerPheLeuHisVal 80
QY 2400 CCTCAGTGGCCAGCACCCCTCTTAGTGTAGAGACTTATTGACCAAGATGGGAGCACAAAT 2459
Db 81 ProAspCysProTyrProLeuLeuGlyArgAspLeuLeuThrLysLeuLysAlaGlnIle 100
QY 2460 TCTTTTCAA---CAAGGGAAACCAAGAGTGTCTCCAAATACAAACCTATCCTGTGTG 2516
Db 101 HisPheGluGlySerGlyAlaGlnValMetGlyProMetGlyGlnProLeuGlnValLeu 120
```

```
QY 2517 ACCTTCCAATTAGATGACGAATATCGACTATACTCTCCCTAGTAAAGCCTGATCAAAAT 2576
Db 121 ThrLeuAsnIleGluAspGluHisArgLeuHisGluThrSerLysGluProAspValSer 140
QY 2577 ATA---CAATCTCTGGTGGAAAGTTCCTCCCAAGCCTGGCGCAAGAACCGCAGGATGGT 2633
Db 141 LeuGlySerThrTrpLeuSerAspPheProGlnAlaTIPAlaGluThrGlyGlyMetGly 160
QY 2634 TTGGCAAGCAAGTTCCTCCCAAGTATTTCACACTGAGCCAGTGCACACCAAGTGTCA 2693
Db 161 LeuAlaValArgGlnAlaProLeuIleIleProLeuLysAlaThrSerThrProValSer 180
QY 2694 GTCAGACAGTACCTCTGAGTAAAGAGCTTCAAGAAGAAATTCGGCCGCATGTCCAAAGA 2753
Db 181 IleLysGlnTyrProMetSerGlnGluAlaArgLeuGlyIleLysProHisIleGlnArg 200
QY 2754 TTAATCCAAACAGCGCATCTCTAGTTCCTGTCCTCAATCTCCCTGGAAATCTCCCTGCTACCG 2813
Db 201 LeuLeuAspGlnGlyIleLeuValProCysGlnSerProTrpAsnThrProLeuLeuPro 220
QY 2814 GTTAGAAAGCCTGGGACTAATGACTATCGACCACTGACAGCTTGAGAGAGTCAATAAA 2873
Db 221 ValLysLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsnLys 240
QY 2874 CGGCTGCAGGATATACACCAACAGTCCCGAACCTTTATTAACCTTGTGTGCTCTCCCA 2933
Db 241 ArgValGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerGlyLeuPro 260
QY 2934 CCCCAACGGAGCTGTATACAGATATTGGACTTAAGAGATGCCTTTCTCTCCCTGAGATTA 2993
Db 261 ProSerHisGlnTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeu 280
QY 2994 CACCCACTAGCCACCACTTTTTCCTTCGAATGGAGAGATCCAGTACGGGAGAAC 3053
Db 281 HisProThrSerGlnProLeuPheAlaPheGluTIPArgAspProGluMetGlyIleSer 300
QY 3054 GGGCAGCTCACCCTGGACCCGACCTGCCCAAGGTTCAAGAACTCCCGACCATCTTTGAC 3113
Db 301 GlyGlnLeuThrTIPThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAsp 320
QY 3114 GAAGCCCTACACAGACCTGGCCAACTTCAGGATCAACACACCTCAGGTGACCTCTCTC 3173
Db 321 GluAlaLeuHisArgAspLeuAlaAspPheArgIleGlnHisProAspLeuLeuLeu 340
QY 3174 CAGTACGTGATGACCTGCTCTGCGGGAGCCACCAACAGGACCTCTTAGAAGGACG 3233
Db 341 GlnTyrValAspAspLeuLeuAlaAlaThrSerGluLeuAspCysGlnGlnGlyThr 360
QY 3234 AAGGCATCTACTGTGGAATTTGCTGACCTAGGCTTACAGAGCTCTGTCTAAGAGGCCAG 3293
Db 361 ArgAlaLeuLeuGlnThrLeuGlyAsnLeuGlyTyrArgAlaSerAlaLysLysAlaGln 380
QY 3294 ATTTGACGAGAGAGAGGTAAACATCTTGGGGTACAGTTTGGGGACGGCGACGATGCTG 3353
Db 381 IleCysGlnLysGlnValLysTyrLeuGlyTyrLeuLeuLysGluGlyGlnArgTipLeu 400
QY 3354 ACGAGGACGCGAAGAAACTGTAGTCCAGATCCGCGCCCAACCAACAGCAACAATG 3413
Db 401 ThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGlnLeu 420
QY 3414 AGAGAGTTTTTGGGACAGCTGGATTTTGACAGCTGTGGATCCCGGGTTTGGACCTTA 3473
Db 421 ArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGluMet 440
QY 3474 GCAGCCCCACTCTACCCGCTAAACCAAGAAAAAGGGGAATTCCTCGGGCTCTCTGACAC 3533
Db 441 AlaAlaProLeuTyrProLeuThrLysThrGlyThrLeuPheAsnTIPGlyProAspGln 460
QY 3534 CAGAGGCATTTGATGCTATCAAAAGGCCCTCTGCTAGCGCACCTGCTGGCCCTCCCT 3593
Db 461 GlnLysAlaTyrGlnGluIleLysGlnAlaLeuThrAlaProAlaLeuGlyLeuPro 480
```







C;Genetics:	
A;Gene: pol	
C;Superfamily: pol polyprotein	
C;Keywords: aspartic proteinase; endonuclease; hydrolase; nucleotidyltransferase; reverse	
F;1-120/Product: proteinase #status predicted <PTN>	
F;3-102/Product: retropepsin #status predicted <RTP>	
F;121-797/Product: RNA-directed DNA polymerase #status predicted <REV>	
F;798-1189/Product: endonuclease #status predicted <EDB>	
F;2/Active site: Asp (shared with dimeric partner) #status predicted	
Alignment Scores:	
Pred. No.:	1.63e-234
Score:	3940.50
Percent Similarity:	76.0%
Best Local Similarity:	63.2%
Query Match:	26.9%
DB:	1
US-10-723-552-3 (1-8132) x GNMVW7 (1-1189)	
Qy	2160 GGGAGAGCGGTTCCGACCCCTCCCGAGCCAGGGTAACCTTTGAAGTGGAGGGGCAA 2219
Db	1 GlyCysGlnGlySerGlyAlaProProGluProArgLeuThrLeuSerValGlyGlyHis 20
Qy	2220 CCAGTTGAGTTCCTGGTTGATACCGGAGCGCAACATTCAGTGTCTACTACAGCCATTAGGA 2279
Db	21 ProThrThrPheLeuValAspThrGlyAlaGlnHisSerValLeuThrLysAlaAsnGly 40
Qy	2280 AAACTAAAGATAAAAACTCTGGGTGATGGTGCCACAGGCGCAACACAGTATCCATGG 2339
Db	41 ProLeuSerSerArgThrSerTrpValGlnGlyAlaThrGlyArgLysMetHisLysTrp 60
Qy	2340 ACTACCCGAAGAACAGTTGACTTGGGAGTGGGACGGGTAAACCCACTGCTTTCTGGTCATA 2399
Db	61 ThrAsnArgArgThrValAsnLeuGlyGlnGlyMetValThrHisSerPheLeuValVal 80
Qy	2400 CCTGAGTGCCAGCACCCCTCTAGGTAGACATTTATGACCAAGATGGGAGCACAAATT 2459
Db	81 ProGluCysProTyrProLeuLeuGlyArgAspLeuLeuThrLysLeuGlyAlaGlnIle 100
Qy	2460 TCTTTT---GAACAAGGAAACCAAGAAGTGTCTGCAAAATRACAAACCTATCAGTGTGTG 2516
Db	101 HisPheSerGluAlaGlyAlaGlnValLeuAspArgAspGlyGlnProIleGlnIleLeu 120
Qy	2517 ACCCTCCAAATTAGATAGCAATATCGACTACTACTCT---CCCTAGTAAAG-----CCT 2567
Db	121 ThrValSerLeuGlnAspGluHisArgLeuPheAspIleProValThrThrSerLeuPro 140
Qy	2568 GATCAAAATATACAATTCTGGTTGSAACAGTTTCCCAAGCTGGCGAGAAACCGCAGGG 2627
Db	141 Asp-----ValTrpLeuGlnAspPheProGlnAlaTrpAlaGluThrGlyGly 156
Qy	2628 ATGGGTTTGGCAAGCAAGTTCCCGCCCAAGATTATTCAACTGAAGGGCCAGTCGCACACCA 2687
Db	157 LeuGlyArgAlaLysCysGlnAlaProIleIleIleAspLeuLysProThrAlaValPro 176
Qy	2688 GTGCAGTCAGACAGTACCCCTTGAGTAAAGAGCTCAAGAGAAATTCGGCCGCGATGTC 2747
Db	177 ValSerIleLysGlnTyrProMetSerLeuGluAlaHisMetGlyIleArgGlnHisIle 196
Qy	2748 CAAGAGTTAATCCACAGGCGATCCTAGTTCCTGCTCCAATCTCCCTGGAACTATCCCGCTG 2807
Db	197 IleLysPheLeuGluLeuGlyValLeuArgProCysArgSerProTrpAsnThrProLeu 216
Qy	2808 CTACCGGTTAGAAAGCCTGGGACTAATGACTATATCGACAGTACAGGACTTCGAGAGAGTC 2867
Db	217 LeuProValLysLysProGlyThrGlnAspTyrArgProValGlnAspLeuArgGluIle 236
Qy	2868 AATAAAGGGTGCAGGATATACACCCACAGTCCCGCAACCCCTTATAACCTCTTGTGTGCT 2927
Db	237 AsnLysArgThrValAspIleHisProThrValProAsnProTyrAsnLeuSerThr 256
Qy	2928 CTCCCACCCCAACGAGTGTGTATACAGTATTGGACTTTAAGGATGCTTCTTCTGCGCTG 2987
Db	257 LeuLysProAspTyrSerTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeu 276
Qy	2988 AGATTACACCCCACTAGCCAAACACATTTTTCCTTGCATGGAGAGATCCAGGTACGGGA 3047
Db	277 ProLeuAlaProGlnSerGlnGluLeuPheAlaPheGluTrpLysAspProGluArgGly 296
Qy	3048 AGAACGGGCGAGCTCACCTCGACCCCGACCTGCCCAAGGGTTCAAGAACTCCCGACCATC 3107
Db	297 IleSerGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeu 316
Qy	3108 TTTGACGAAGCCCTTACACAGAGACCTGGCCCAACTTTCAGGATCCAAACACCCCTCAGGTGACC 3167
Db	317 PheAspGluAlaLeuHisArgAspLeuThrAspPheArgThrGlnHisProGluValThr 336
Qy	3168 CTCCTCCAGTACGTGGATGACCTGCTTCTCGGGGAGCCACCAACAGGACTGCTGTAGAA 3227
Db	337 LeuLeuGlnTyrValAspAspLeuLeuAlaProThrLysLysAlaCysThrGln 356
Qy	3228 GGCACGAAGCAGCTACTGCTGGAAATGCTCTGACCTAGGCTACAGAGCTCTGTCTAAGAAG 3287
Db	357 GlyThrArgHisLeuLeuGlnGluLeuGlyGluLysGlyTyrArgAlaSerAlaLysLys 376
Qy	3288 GCCCAGATTTGCAGGAGAGGTAAACATCTTGGGGTACAGTTTTCGGGAGCGGCACCGA 3347
Db	377 AlaGlnIleCysGlnThrLysValThrTyrLeuGlyTyrIleLeuSerGluGlyLysArg 396
Qy	3348 TGCTGACGAGGCGACGGAAGAAAACCTGTAGTCCAGATACCGGCCCCCAACACAGCCAAA 3407
Db	397 TrpLeuThrProGlyArgIleGluThrValAlaArgIleProProArgAsnProArg 416
Qy	3408 CAAATGAGACAGTTTTGGGAGCAGCTGGATTTTGCAGACTGTGATCCCGGGGTTTGGC 3467
Db	417 GluValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAla 436
Qy	3468 ACCTTAGCAGCCCACTCTACCCGCTAAACCAAGAAAGGGGAATTTCTCTGGGCTCCT 3527
Db	437 GluLeuAlaAlaProLeuTyrAlaLeuThrLysGluSerThrProPheThrTrpGlnThr 456
Qy	3528 GAGCACAGAGGATTTGATGATCAAAAAGCCCTGCTGAGCGCACCTGCTCTGCGCC 3587
Db	457 GluHisGlnLeuAlaPheGluAlaLeuLysLysAlaLeuLeuSerAlaProAlaLeuGly 476
Qy	3588 CTCCTGACCTAACTAAACCTTTACCCCTTTATGTGATGAGCGTAAAGGATGAGCCGG 3647
Db	477 LeuProAspThrSerLysProPheThrLeuPheLeuAspGluArgGlnGlyIleAlaLys 496
Qy	3648 GAGTTTTTAACCCAAACCTTAGGACCATGAGAGAACCTGTCGCTACCTGTCTCAAGAAG 3707
Db	497 GlyValLeuThrGlnLysLeuGlyProTrpLysArgProValAlaTyrLeuSerLysLys 516
Qy	3708 CTGATCCTGTAGCCAGTGTGGCCCATATGCTGAAGCTATCGAGCTGTGGCCATA 3767
Db	517 LeuAspProValAlaAlaGlyTyrProProCysLeuArgIleMetAlaAlaThrAlaMet 536
Qy	3768 CTGCTCAAGCAGCTGACAAATTCAGCTTTGGGACAGATATACCTGTAATAGCCCCCAT 3827
Db	537 LeuValLysAspSerAlaLysLeuThrLeuGlyGlnProLeuThrValIleThrProHis 556
Qy	3828 GCATTGGAGAACATCGTTTCGGCAGCCCCCAGACCGATGGATGACCAACGCCCGCATGACC 3887
Db	557 ThrLeuGluAlaIleValArgGlnProProAspArgTrpIleThrAsnAlaArgLeuThr 576
Qy	3888 CACTATCAAGCCTGCTTCTC---ACAGAGGGGTACGTTTCGCTCCACAGCCGCTCTC 3944
Db	577 HisTyrGlnAlaLeuLeuLeuAspThrAspArgValGlnPheGlyProProValThrLeu 596
Qy	3945 AACCTCGCACTCTTCTGCTGAGAGACTGATGAACACAGTACTGATGATGTCATCA 4004
Db	597 AsnProAlaThrLeuLeuProValProGluAsnGlnProSerProHisAspCysArgGln 616
Qy	4005 CTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAGAA 4064

Db 617 ValLeuAlaGluThrHisGlyThrArgGluAspLeuLysAspGlnGluLeuProAspAla 636  
 QY 4065 GTGCTAACCTGGTTCTACTGACGAGCAGCTATGTGTGAAGTAAAGAGGATGGCTGGG 4124  
 Db 637 AspHisThrTrpTyThrAspGlySerSerTyrLeuAspSerGlyThrArgAlaGly 656  
 QY 4125 GCGGGGTGGTGGAGGAGCCGACCATCTGGCCAGAGCTCGCCGGAGGAACTTCA 4184  
 Db 657 AlaAlaValAlaAspGlyHisAsnThrIleTrpAlaGlnSerLeuProProGlyThrSer 676  
 QY 4185 GCACAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGGCGTGGCCGAGGGAATCC 4244  
 Db 677 AlaGlnLysAlaGluLeuIleAlaLeuThrLysAlaLeuGluLeuSerLysGlyLys 696  
 QY 4245 ATAAACATTATACGACAGCAGGTATGCCCTTTGCGACTGCGACACGCTACATGGGCCCATC 4304  
 Db 697 AlaAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisThrHisGlySerIle 716  
 QY 4305 TATAAACAAAGGGGTGGTCTTACTCTACGAGGAGGGAATAAAGAACAAAGAGGAAT 4364  
 Db 717 TyrGluArgArgGlyLeuLeuThrSerGluGlyLysGluIleLysAsnLysAlaGluIle 736  
 QY 4365 CTAAGCCTATTAGAAGCGCTACATTTACCAAAAAGGCTAGCTATTATACACTGTCCTGGA 4424  
 Db 737 IleAlaLeuLeuLysAlaLeuPheLeuProGlnGluValAlaIleIleHisCysProGly 756  
 QY 4425 CATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGAGTGGCTGACCGGGTTGCCAAG 4484  
 Db 757 HisGlnLysGlyGlnAspProValAlaValAlaGlyAsnArgGlnAlaAspArgValAlaArg 776  
 QY 4485 CAGCAGCCAGGGTGTAACTCTCTCCCTATATAGAAATGCCAAAGCCCGACAGACCC 4544  
 Db 777 GlnAlaAla-----MetalGluValLeuThrLeuAlaThrGluProAspAsn 792  
 QY 4545 -----AGACGACATCACCTCAGACACTGGCAAGAGATAAAGAG 4586  
 Db 793 ThrSerHisIleThrIleGluHisThrTyrThrSerGluAspGlnGluAlaAlaArgAla 812  
 QY 4587 ATAGACAGTTCTCTGAGACTCCGAGAGGACCTGCTATACCTCAGATGGGAAGGAAATC 4646  
 Db 813 IleGlyAla---ThrGluAsnLysAspThrArgAsnTrpGluLysGluGlyLysIleVal 831  
 QY 4647 CTCGCCCAAGAGAGGTTAGATATGTCCACAGATATATCTGCTAACCCACCTAGGA 4706  
 Db 832 LeuProGlnLysGluAlaLeuAlaMetIleGlnMetHisAlaTrpThrHisLeuGly 851  
 QY 4707 ACTAAACACTGCGAGCTGGTCAGAACATCCCTTATCATCTTCTGAGCTACACGGA 4766  
 Db 852 AsnArgLysLeuLysLeuIleGluLysThrAspPheLeuIleProArgAlaSerThr 871  
 QY 4767 GTGGCTGACTCGGTGCTCAACATTTGTGCTCCCTGCGAGCTGTTAATGCTAATCTCTCC 4826  
 Db 872 LeuIleGluGlnValThrSerAlaCysLeuValCysGlnGlnValAsnAlaGlyAlaThr 891  
 QY 4827 AGAATGCTCCAGGAGAGACTAAGGGAGCCAGCCAGGCGCTCACTGGGAAGTGGAC 4886  
 Db 892 ArgValProAlaGlyLysArgThrArgGlyAsnArgProGlyValTyrTrpGluIleAsp 911  
 QY 4887 TTCACGTAGTAAAGCCGGCTAATACGGGAACAAATACCTATTGCTTTTGTGACACCC 4946  
 Db 912 PheThrGluValLysProHisTyrAlaGlyTyrLysTyrLeuLeuValPheValAspThr 931  
 QY 4947 TTTTCAGGATGGGTAGAGGCTTATCTCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 5006  
 Db 932 PheSerGlyTrpValGluAlaPheProThrArgGlnGluThrAlaHisIleValAlaLys 951  
 QY 5007 AAATACTGGAAGAAATTTTCCAGATTGGGAATACCTAAGGTAATAGGGTCAGACAAT 5066  
 Db 952 LysIleLeuGluGluIlePheProArgPheGlyLeuProLysValIleGlySerAspAsn 971  
 QY 5067 GGTCCAGCTTTTGTCCCGAGGTAACTCAGGAGCTGCCCAAGATATTGGGGATTGATGG 5126  
 Db 972 GlyProAlaPheValSerGlnValSerGlnGlyLeuAlaArgIleLeuGlyIleAsnTrp 991

QY 5127 AAAGTGCATTGTGCATACAGACCCCAAGCTCAGACAGGTAGAGAGTGAATAGAACCC 5186  
 Db 992 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 1011  
 QY 5187 ATTAAGAGACCTTACTTAATTCAGCCGCGAGACTGGCTGCTTAATGATTGGATAGCTCTC 5246  
 Db 1012 IleLysGluThrLeuThrLysLeuGluThrLeuGluThrGlyLeuLysAspTrpArgArgLeu 1031  
 QY 5247 CTCGCCCTTTGTGCTTTTAGGGTTAGGAACACCCCTGGACAGCTTGGGCTGACCCCTAT 5306  
 Db 1032 LeuSerLeuAlaLeuAlaArgAlaArgAsnThrProAsnArgPheGlyLeuThrProTyr 1051  
 QY 5307 GAATTACTCTACCGGGGACCCCTTCTGTAGAAATTCCTCTGTACATAGTGTGCAC 5366  
 Db 1052 GluIleLeuTyrGlyGlyProProLeuSerThrLeuLeuAsnSerPheSerProSer 1071  
 QY 5367 GTCTGCTTCTCCAGCCTTTGCTCTAGGCTCAAGGCACTTGTAGTGGGTGAGACAACGA 5426  
 Db 1072 AsnSerLysThrAsp---LeuGlnAlaArgLeuLysGlyLeuGlnAlaValGlnAlaGln 1090  
 QY 5427 GCGTGGAGGCAACTCCGGGAGGCTACTCAGGAGGAGGAGACTTGCAGATCCCAACATCGT 5486  
 Db 1091 IleTrpAlaProLeuAlaGluLeuTyrArgProGlyHisSer---GlnThrSerHisPro 1109  
 QY 5487 TTCCTAAGTGGGATTCAGTCTTACGTAGACGCCACCGTGCAGAAACCTCGAGACTCGG 5546  
 Db 1110 PheGlnValGlyAspSerValTyrValArgArgHisArgSerGlnGlyLeuGluProArg 1129  
 QY 5547 TGAAGGGCCCTTATCTCGTACTTTTGACACACCAACCGCTGTGAAGTCGAGGAATC 5606  
 Db 1130 TrpLysGlyProTyrIleValLeuLeuThrThrProThrAlaIleLysValAspGlyIle 1149  
 QY 5607 TCACCTGGATCCATCCATCCCGCTTAAACCGCGCCA-----CCTCCCGAT 5654  
 Db 1150 AlaThrTrpIleHisAlaSerHisAlaLysAlaAlaProGlyThrProGlyProThrSer 1169  
 QY 5655 TCGGGG---TGGAAAGCCGAAAGACTGAAATTCCTTAAAGTTCGCTCCATCCG 5708  
 Db 1170 SerGlyThrTrpArgLeuArgArgSerGluAspProLeuLysIleArgLeuSerArg 1188  
 RESULT 9  
 GNMVCE  
 pol polyprotein - feline endogenous virus ECE1 (fragment)  
 N;Contains: endonuclease (EC 3.1.-.-); RNA-directed DNA polymerase (EC 2.7.7.49)  
 C;Species: feline endogenous virus ECE1  
 A;Note: host Felis silvestris catus (domestic cat)  
 C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
 C;Accession: S12813; S12814  
 R;Moehring, R.  
 submitted to the EMBL Data Library, February 1990  
 A;Reference number: S12812  
 A;Accession: S12813  
 A;Molecule type: DNA  
 A;Residues: 1-1046 <MO>  
 A;Cross-references: UNIPROT:P31792; UNIPARC:UPI0000131ED7; EMBL:X51929  
 C;Gene: pol  
 C;Superfamily: pol polyprotein  
 C;Keywords: hydrolase; nucleotidyltransferase; polyprotein  
 F;1-647/Product: RNA-directed DNA polymerase (fragment) #status predicted <RTP>  
 F;648-1046/Product: endonuclease #status predicted <EDC>

Alignment Scores:  
 Pred. No.: 1.18e-209 Length: 1046  
 Score: 3536.50 Matches: 676  
 Percent Similarity: 76.4% Conservative: 129  
 Best Local Similarity: 64.1% Mismatches: 228  
 Query Match: 24.2% Indels: 21  
 DB: 1 Gaps: 8

US-10-723-552-3 (1-8132) x GNMVCE (1-1046)







Db	718	LysThrAspPheLeuIleProLysAlaGlyThrLeuIleGluGlnValThrSerAlaCys	737
QY	4794	GTGCCCTGCCAGCTGGTAAATGCTAAATCTCTCCAGAATGCCTCCAGGGAAGAGACTAAGG	4853
Db	738	LysValCysGlnGlnValAsnAlaGlyAlaThrArgValProGluGlyLysArgThrArg	757
QY	4854	GGAAAGCACCCAGGCGCTCACTGGGAAGTGCACCTCACTGAGGTAAAGCCGGCTAAATAC	4913
Db	758	GlyAsnArgProGlyValTyrTrpGluIleAspPheThrGluValLysProHisTyrAla	777
QY	4914	GGAAACAAATACCTATTGGTTTTGTATAGACACCTTTTCAGGATGGGTAGAGCTTATCCT	4973
Db	778	GlyTyrLysTyrLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaTyrPro	797
QY	4974	ACTAAGAAAGAGACTTCAACCGTCGTGGCTAAAAAATACATGGAAGAAATTTTTCCAAGA	5033
Db	798	ThrArgGlnGluThrAlaHisMetValAlaLysLysIleLeuGluGluIlePheProArg	817
QY	5034	TTTGGATACCTAAGGTAATAGGTCAGACAAATGGTCCAGCTTTTGTGGCCAGGTAAAT	5093
Db	818	PheGlyLeuProLysValIleGlySerAspAsnGlyProAlaPheValSerGlnValSer	837
QY	5094	CAGGCACTGGCCAAATATTGGGATTTGATGGAAATGCAATTCGCATACAGACCCCAA	5153
Db	838	GlnGlyLeuAlaArgThrLeuGlyIleAsnTrpLysLeuHisCysAlaTyrArgProGln	857
QY	5154	AGTCAGGACAGGTAGAGAGATGAATAGAACCAATTAAGAGACCCCTTACTAAATTCACC	5213
Db	858	SerSerGlyGlnValGluValArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThr	877
QY	5214	CGCGAGACTCGCGTTAATGATTGGATAGCTCTCTGCGCCCTTTGTGCTTTTGGGTTAGG	5273
Db	878	LeuGluThrGlyLeuLysAspTrpArgLeuLeuLeuSerLeuAlaLeuLeuArgAlaArg	897
QY	5274	AACACCCCTCGACATTTGGCGTCAGCCCTTATGAATTACTCTACGGGGGACCCCCCA	5333
Db	898	AsnThrProAsnArgPheGlyLeuThrProTyrGluIleLeuTyrGlyGlyProProPro	917
QY	5334	TTGGTAGAAATATGCTTCTGTACATAGTGTGACGTGTCTGCTTCCACGCTT-----	5384
Db	918	LeuSerThrLeuLeuAsnSerPheSerPro-----SerAspProLysThrAsp	933
QY	5385	TTGTTCTCTAGGCTCAAGGCACCTTGAGTGGGTGAGACAAACGAGCGTCGAGGCACTCCGG	5444
Db	934	LeuGlnAlaArgLeuLysGlyLeuGlnAlaValGlnAlaGlnIleTrpThrProLeuAla	953
QY	5445	GAGCGCTACTTCAGGAGGAGAGACTTCGACATCCACATCGTTTCCAAATGGGAGATTCA	5504
Db	954	GluLeuTyrArgProGlyHisPro--GlnThrSerTyrProPheGlnValGlyAspSer	972
QY	5505	GTCTAGCTTTAGACGCCACCGTCGAGGAAACCTCGAGACTCGGTGGAGGCGCTTATCTC	5564
Db	973	ValTyrValArgTrpHisArgSerGlnGlyLeuGluProArgTrpLysGlyProTyrIle	992
QY	5565	GTACTTTTGACACCAACGGCTGTGAAGTGAAGGAATCTCCACCTCGATCCATGCA	5624
Db	993	ValLeuLeuThrThrProThrAlaIleLysValAspGlyIleAlaIleTrpIleHisAla	1012
QY	5625	TCCCACTGTTAAA-----CCGGCGCCACTCCCGATTCCGGGTGGAAAGCC	5669
Db	1013	SerHisAlaLysAlaAlaProLysThrProGlyProGluThrProLysThrTrpLysLeu	1032
QY	5670	GAAGAAGACTGAAATCCCTTAAGCTTTCGCTCCATCGCGTG	5711
Db	1033	HisArgSerGluAsnProLeuLysIleArgLeuSerArgVal	1046

## RESULT 10

GNVWIK

pol polyprotein - AKR murine leukemia virus (fragment)

C/Species: AKR murine leukemia virus

C; Date: 05-Apr-1983 #sequence revision 05-Apr-1983 #text change 09-Jul-2004

C;Accession: A03958

R; Herr, W.; Corbin, V.; Gilbert, W.

Nucleic Acids Res. 10, 6931-6944, 1982  
A;Title: Nucleotide sequence of the 3' half of AKV.

**A;Title:** Nucleotide sequence of the 3' half of AKV.

A;Reference number: A93448; MUID:83090450; PMID:6294621

A;Accession: A03958

A;Molecule type: DNA

A;Residues: 1-843 &lt;HER&gt;

A; Cross-references: UNIPROT: P03357; UNIPARC: UPI0000131F01

C;Genetics:

A;Gene: pol

C:superfamily: pol polyprotein

C: Keywords: polyprotein

Alignment Scores:		
Pred. No.:	1.86e-154	Length:
Score:	2339.50	Matches:
Percent Similarity:	75.5%	Conservative:
Best Local Similarity:	59.7%	Mismatches:
Query Match:	18.0%	Indels:
DB:	1	Gaps:
		843
		504
		133
		194
		13
		8

US-10-723-552-3 (1-8132) x GNVWK (1-843)

[illegible]

```
QY 3993 GATTGCCATCACTATTGTTGAGGAGACTGGGGTCCGAAAGGACCTTTACAGACATACCG 4052
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 262 AspCysLeuGluIleLeuAlaGluThrHisGlyThrArgProAspLeuThrAspGlnPro 281
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4053 CTGACTGGAGAAGTCTAACCTGGTTCTACTACGGAAGCAGCTATGTGGTGAAGTAAG 4112
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 282 IleProAspAlaAspHisThrTrpTyThrAspGlySerSerPheLeuGlnGluGlyGln 301
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4113 AGGATGGTGGGGCGGGTGGTGGAGCGGACCCGACGATCTGGCGCAGCGCTGCCG 4172
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 302 ArgIysAlaGlyAlaAlaValThrThrGluThrGluValIleTrpAlaArgAlaLeuPro 321
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4173 GAAGGAACTTCAGCAAAAGCGCTGAGCTCATGGCCCTCAGCAAGCTTTGGCGCTGCC 4232
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 322 AlaGlyThrSerAlaGlnArgAlaGluIleAlaLeuThrGlnAlaLeuIysMetAla 341
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4233 GAAGGAAATCCATAAACATTATACGGACAGCAGGTATGCTTTGCCACTGCACACGTA 4292
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 342 GluGlyIysArgLeuAsnValTyThrAspSerArgTyAlaPheAlaThrAlaHisIle 361
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4293 CATGGGCCATCTATAAACAAGGGGTGTTACCTCAGCAGGGAGGGAATAAAGAAC 4352
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 362 HisGlyGluIleTyArgArgGlyLeuLeuThrSerGluGlyArgGluIleIysAsn 381
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4353 AAAGGGAATCTAAGCCTATTAGAACCGCTACATTTACCAAAAGCGCTAGCTATTATA 4412
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 382 LysSerGluIleLeuAlaLeuIysAlaLeuPheLeuProLysArgLeuSerIleIle 401
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4413 CACTGTCTCGACATCAGAAAGCTAAAGATCTCATATCAGAGGAAACACAGATGGCTGAC 4472
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 402 HisCysLeuGlyHisGlnIysGlyAspSerAlaGluAlaArgGlyAsnArgLeuAlaAsp 421
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4473 CGGGTTCGCAAGCAGCGACGCC-----CAGGGTGTTAACTTCTGCTCATATA 4520
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 422 GlnAlaAlaArgGluAlaAlaIleIysThrProProAspThrSerThrLeuLeuIleGlu 441
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4521 GAATGCCCCAAGCCCCAGACCCAGACGACGACGACCTAGAACGCTGCCAAGAGATA 4580
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 442 AspSerThrProTyThrProAlaTyPheHisTyThrGluThrAspLeuLysLeu 461
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4581 AAAAAGATAGACCACTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAG 4640
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 462 ArgGluLeuGlyAlaThrTyAsnGlnSerIysGly---TyTrpValPheGlnGlyLys 480
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4641 GAAATCTGCCCAACAAGAGGGTTAGATATGTCACACAGATACATGCTCTAACCCAC 4700
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 ProValMetProAspGlnPheValPheGluLeuLeuAspSerLeuHisArgLeuThrHis 500
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4701 CTAGGAACTAACACCTGCAGCAGTTGTC-----AGAACATCCCTTATCATGTTCTG 4754
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 501 LeuGlyTyGlnLysMetLeuAlaLeuLeuAspArgGlyGluSerProTyTyMetLeu 520
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4755 AGGCTACGAGAGTGGCTGACTCGGTGTCGCTCAAAATTTGTGTCCTGCCAGCTGTTAAT 4814
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 521 AsnArgAspLysThrLeuGlnTyValAlaAspSerCysThrValCysAlaGlnValAsn 540
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4815 GCTAATCTTCCAGATGCTCCAGGGAAGAGACTAAGGGAAGCCACCCAGGCGCTCAC 4874
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 AlaSerIysAlaLysIleGlyAlaGlyValArgValArgGlyHisArgProGlySerHis 560
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4875 TGGGAAGTGGCTTCACTAGGTAAAGCCGCTAAATACGGAACAAATACCTATTGGTT 4934
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 561 TrpGluIleAspPheThrGluValLysProGlyLeuTyGlyTyLysTyLeuLeuVal 580
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4935 TTTGTAGACACTTTTACAGTAGGTGCTATCTCTACTAAGAAAGACTTCAACC 4994
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 581 PheValAspThrPheSerGlyTrpValGluAlaPheProThrLysArgGluThrAlaArg 600
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4995 GTGTGGCTAAAAATACTCGAAGAAATTTTCCAAGATTTGGAATACCTTAAGGTATA 5054
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 ValValSerIysLysLeuLeuGluIlePheProArgPheGlyMetProGlnValLeu 620
```

```
QY 5055 GGGTCAGACAATGGTCCAGCTTTTGTGCCCCAGGTAAAGTCAGGACTGGCCAAAGATATTG 5114
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 621 GlySerAspAsnGlyProAlaPheThrSerGlnValSerGlnSerValAlaAspLeuLeu 640
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5115 GGGATTGATTGGAAACATGTCATTGTGCATACAGACCCCAAGCTCAGGACAGTAGAGG 5174
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 641 GlyIleAsp---LysLeuHisCysAlaTyArgProGlnSerSerGlyGlnValGluArg 659
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5175 ATGAATAGAACCATTAAGAGACCCCTTACTAATATTACCGCGGAGACTGGCGTTAATCAT 5234
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 660 MetAsnArgThrIleLysGluThrLeuThrLysLeuThrLeuAlaAlaGlyThrArgAsp 679
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5235 TGGATAGCTCTCCGCGCTTTGTGCTTTTAGGTAGGTAAGACACCCCTGGACAGTTTCGG 5294
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 680 TrpValLeuLeuLeuProLeuAlaLeuTyArgAlaArgAsnThrProGlyProHisGly 699
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5295 CTGACCCCTATGTAATTACTCTACGGGGGACCCCCCTTGTGTAGAAAATTCGTTCTGTA 5354
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 700 LeuThrProTyGluIleLeuTyArgAlaProProLeuValAsnPheHisAspPro 719
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5355 CATAGTGTGACGTGCTGCTTTCCAGCGCTTTCTTAGGCTCAAGGCACTTGAGTGG 5414
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 720 AspMetSerGluLeuThrAsnSerProSerLeuGlnAlaHisLeuGlnAlaLeuGlnThr 739
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5415 GTGAGACAACGAGCGTGGAGCAACTCCGGGAGGCTACTCAGGAGGAGGACTTGCAG 5474
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 740 ValGlnArgGluIleTyrLysProLeuAlaGluAlaTyArgAspGlnLeuAspGlnPro 759
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5475 ---ATCCCATCATGTTTCCAAAGTGGGAGATTTCAGTCTACGTTAGACGCCACCGTCAGGA 5531
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5532 AACCTCAGACTCGGTGGAGGGCCCTTATCTCGTACTTTTGACACACCAACCGCTGTG 5591
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 780 AsnLeuGluProAspGlyProTyThrValLeuLeuThrThrProThrAlaLeu 799
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5592 AAAGTCGAAGAACTCCACCTGGATCCATCCATCCACAGCTTAAACCGGCG- - - - -CCA 5645
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 800 LysValAspGlyIleSerAlaTrpIleHisAlaHisValLysAlaAlaThrThrPro 819
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5646 CCTCCCGATTGGGGTGGAAAGCCGAAAGCTGAAATCCCTTAAAGCTTCGCTCCAT 5705
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 820 ProIleLysProSerTrpArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThr 839
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5706 CGCGTGGTTCT 5717
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 840 ArgGlyAlaPro 843
Db ||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 11
A46311
pol polyprotein - Moloney murine leukemia virus (strain 3-1R) (fragment)
C:Species: Moloney murine leukemia virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 31-Dec-2004
C:Accession: A46311
R:Martinelli, S.C.; Goff, S.P.
Virology 174, 135-144, 1990
A:Title: Rapid reversion of a deletion mutation in Moloney murine leukemia virus by recombination
A:Reference number: A46311; MUID:90101369; PMID:2294637
A:Accession: A46311
A:Molecule type: DNA
A:Residues: 1-559 <MAR>
A:Cross-references: UNIPROT:Q957M1; UNIPROT:Q83362; UNIPROT:Q92808; UNIPROT:Q83371; UNIPROT:Q83372
A:Note: readthrough of the terminator TAG occurs between the codons GAC for residue 228-229
C:Genetics:
A:Gene: pol
C:Keywords: polyprotein
Alignment Scores:
Pred. No.: 1.14e-106 Length: 559
Score: 1862.50 Matches: 362
Percent Similarity: 74.8% Conservative: 69
Best Local Similarity: 62.8% Mismatches: 108
Query Match: 12.7% Indels: 37
```

DB: 2 Gaps: 9

US-10-723-552-3 (1-8132) x A46311 (1-559)

1476 TTTCCCTTAACCTCGCCCGGTTGGGACTACACAGCGCTGAAGGTAGGAGAGCTTGAA 1535  
 2 PheProLeuGluArgProAspTrpAspTrpThrGlnAlaGlyArgAsnHisLeuVal 21

1536 ATCTATCCCGAGGCTCTGGTGGCGGCTCCGGGCGCTCAAGAGCGGCCCACTAAATTTG 1595  
 22 HisTyrArgGlnLeuLeuAlaGlyLeuGlnAsnAlaGlyArgSerProThrAsnLeu 41

1596 CTAAGGTAAAGAGATGATCGAGGACCGAATGAACCCCTCTGTTTCTTGAGAGG 1655  
 42 AlaLysValLysGlyIleThrGlnGlyProAsnGluSerProSerAlaPheLeuGluArg 61

1656 CTCCTTGAAGCCTTCAGCGGTACACCCCTTTTATCCACCTCAGAGGCCCAAGGCC 1715  
 62 LeuLysGluAlaTyrArgTyrThrProTyrAspProGluAspProGlyGlnThr 81

1716 TCAGTGGCTTTGGCCTTTATAGGACAGTCAGCCTTGGATATTAGAAAGAGCTTCAGAGA 1775  
 82 AsnValSerMetSerPheIleTrpGlnSerAlaProAspIleGlyArgLysLeuGlyArg 101

1776 CTGGAAGGTTACAGGAGCTGAGTTACGTATCTAGTGAAGGAGGAGAGAAATATAT 1835  
 102 LeuGluAspLeuLysSerLysThrLeuGlyAspLeuValArgGluAlaGluLysIlePhe 121

1836 TACAAAAGGAGACAGAGAGAGAGGAAACAAAGAAAGAGAGAGAGAGAGAGAGG 1895  
 122 AsnLysArgGluThrProGluGluArgGluGluArgIleArgArgGluThrGluGluLys 141

1896 GAGGAAAGA-----CGTAATAAAGCGCAAG 1922  
 142 GluGluArgArgThrValAspGluGlnLysGluLysGluArgAspArgArgHis 161

1923 AAGAAATTTGACTAAGATCTTGGCTGCGAGTGTGAAGGAAAGCAATACGGAAGAGAG 1982  
 162 ArgGluMetSerLysLeuLeuAlaThrValValIleGly-----GlnGlu 176

1983 AGAGATTTAGGAAATTAGTCCAGGCCCTAGACAGTCAGGAACCTGGGCGMATAGACC 2042  
 177 GlnAsp-----ArgGlnGluGlyGlu---ArgLysArgPro 187

2043 CCACCTGCACAGGACCAATGTGCATATTGTAAGAAAGAGACACTGGGCAAGAACTGC 2102  
 188 GlnLeuAspLysAspGlnCysAlaTyrCysLysGluLysGlyHisTrpAlaLysAspCys 207

2103 CCCAAGAG-----GGAAACAAAGGACCAAG-----ATCCTAGCTCTAGAA 2144  
 208 ProLysLysProArgGlyProArgGlyProArgGlyProArgProGlnThrSerLeuThrLeuGly 227

2145 GAAGATAAGATTAGGAGACGCGGTTCCGACCCCTCCCGAGCCAGCGGTAACTTTG 2204  
 228 Asp-----GlyGlyGlnGlyGlnAspProProGluProArgIleThrLeu 243

2205 AAGGTGAGGGGCAACCAAGTTGAGTCTCTGTTGATACCGGAGCGAAACATTTCAGTCTA 2264  
 244 LysValGlyGlyGlnProValThrPheLeuValAspThrGlyAlaGlnHisSerValLeu 263

2265 CTACAGCCATTAGGAAACTAAAGATAAAAAATCTCGGTGATGGTGCACAGGGCAA 2324  
 264 ThrGlnAsnProGlyProLeuSerAspLysSerAlaTrpValGlnGlyAlaThrGlyGly 283

2325 CAACAGTATCATGACTACCGAGAACAGTTGACTTTGGAGTGGGACGGGTAAACCCAC 2384  
 284 LysArgTyrArgTrpThrAspArgLysValHisLeuAlaThrGlyLysValThrHis 303

2385 TCGTTTCTGCTGATCTGAGTCCAGCAGCCCTCTTAGTAGAGACTTATGACCAAG 2444  
 304 SerPheLeuHisValProAspCysProTyrProLeuLeuGlyArgAspLeuLeuThrLys 323

2445 ATGGGAGCACAAATTTCTTTTGAA---CAAGGGAAACAGAAAGTGTCTGCAAAATAACAA 2501

Db 324 LeuLysAlaGlnIleHisPheGluGlySerGlyAlaGlnValValGlyProMetGlyGln 343  
 2502 CCTATCACTGTGTGACCCCTCAATTAGATGACCAATATCCGATATCTCTCCCTAGTA 2561  
 344 ProLeuGlnValLeuThrLeuAsnIleGluAspGluTyrArgLeuHisGluThrSerLys 363  
 2562 AAGCCTGATCAAAATATATCAATTC---TGGTTGAAACAGTTTCCCAAGCCTGGGCAGAA 2618  
 364 GluProAspValSerLeuGlyPheThrTrpLeuSerAspPheProGlnAlaTrpAlaGlu 383  
 2619 ACCCGAGGATGGTTTGGCAAGCAAGTTCCCCCAACAAGTTATCAACTCAAGGCCAGT 2678  
 384 SerGlyGlyMetGlyLeuAlaValArgGlnAlaProLeuIleIleProLeuLysAlaThr 403  
 2679 GCCACACCATGTCAGTCAGACAGTACCCCTTGGTAAAGAGCTCAAGAGGAATTCGG 2738  
 404 SerThrProValSerIleLysGlnTyrProMetSerGlnGluAlaIleArgLeuGlyIleLys 423  
 2739 CCGCATGTCCAAAGATTAAATCAACAGGCGATCTAGTTCCTGCTCCCAATCTCCCTGGAAT 2798  
 424 ProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCysGlnSerProTrpAsn 443  
 2799 ACTCCCTGCTACCGGTTAGAAAGCCTGGGCACTAATCACTATCATCCAGTACAGGACTTG 2858  
 444 ThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArgProValGlnAspLeu 463  
 2859 AGAGAGTCAATAAACCAGGTCAGGATATACACCAACAGTCCCGAACCTTATTAACCTC 2918  
 464 ArgGluValAsnLysArgValGluAspIleHisProThrValProAsnProTyrAsnLeu 483  
 2919 TTGTGTCTCTCCACCCCAACAGGAGCTGTGTATACAGTATTGGACTTTAAAGGATGCCTTC 2978  
 484 LeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAspLeuLysAspAlaPhe 503  
 2979 TTCTGCTGAGATTACACCCACTAGCCAAACCACTTTTTCCTTCGAATGGAGATGCCA 3038  
 504 PheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGluTrpArgAspPro 523  
 3039 GGTACGGAGAACACCGGCGAGCTCACCTGACCGGCTGCGCCGAGGGTTCAGAACTCC 3098  
 524 GluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSer 543  
 3099 CCGACCATCTTTTCACGAAGCCCTACACAGAGACTGGCCCACTTCAGG 3146  
 544 ProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPheArg 559

RESULT 12  
 S04842  
 pol polyprotein - simian sarcoma-associated virus (fragment)  
 N;Contains: endonuclease; ribonuclease H (EC 3.1.26.4)  
 C;Species: simian sarcoma-associated virus, SSAN  
 C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 07-Feb-1997  
 C;Accession: S04842  
 R;Brack-Werner, R.; Werner, T.; Leib-Moesch, C.; Kehlmann, R.; Erfle, V.  
 Nucleic Acids Res. 17, 5382, 1989  
 A;Title: Primary structure of the SSAN tether-RNase H endonuclease (pol) region deleted  
 A;Reference number: S04842; MUID:89345098; PMID:2548158  
 A;Molecule type: mRNA  
 A;Residues: 1-476 <BRA>  
 A;Cross-references: UNIPARC:UPI00001785C4; EMBL:X15311  
 C;Genetics:  
 A;Gene: pol  
 C;Superfamily: pol polyprotein  
 C;Keywords: hydrolase; polyprotein  
 F;1-205/Product: tether (fragment) #status predicted <ET>  
 F;206-361/Product: ribonuclease H #status predicted <RNH>  
 F;362-476/Product: endonuclease (fragment) #status predicted <END>

Alignment Scores: 2,96e-98 Length: 476  
 Pred. No.: 1725.50 Matches: 335  
 Score:

Percent Similarity: 76.6%		Conservative: 49
Best Local Similarity: 66.9%		Mismatches: 92
Query Match: 11.8%		Indels: 25
DB: 2		Gaps: 4
US-10-723-552-3 (1-8132) x S04842 (1-476)		
QY	3441	TGCAGACTGTCGATCCCGGGTTCGACCTTAGCAGCCCACTCTACCGCTACCAAA 3500
Db	1	CysargLeuTrpIleProGlyPheAlaSerLeuAlaProLeuTyrProLeuThrLys 20
QY	3501	GAATAAGGGGAATTCCTCCGTGGCTCCTGAGCACCAGAGGCAATTGTGCTATCAAAAG 3560
Db	21	GluserIleProPheIleTrpThrGluGluHisGlnLysAlaPheAspArgIleLysGlu 40
QY	3561	GCCCTGTGAGCGCAGCTGCTCTGGCCCTCCTGAGCACTAAACCTTTACCTTTAT 3620
Db	41	AlaLeuLeuSerAlaProAlaLeuAlaLeuProAspLeuAlaLysProPheThrLeuTyr 60
QY	3621	GTGATGAGCGTAAGGAGTAGCCGGGAGCTTTTAACCCAAACCTAGGACCATGGGA 3680
Db	61	ValAspGluArgAlaGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArg 80
QY	3681	AGACTGTGCTGCTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGC 3740
Db	81	ArgProValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProThrCys 100
QY	3741	CTGAAGGTATCGCAGCTGTGGCCATCTGCTCAAGACGCTGACAAATTGACTTTGGGA 3800
Db	101	LeuLysAlaValAlaAlaValAlaLeuLeuLysAspAlaAspLysLeuThrLeuGly 120
QY	3801	CAGATATACTGTAATAGCCCCCATCATCTGATGGAGAACATCGTTCGGCAGCCCCAGAC 3860
Db	121	GlnAsnValThrValIleAlaSerHisSerLeuGluSerIleValArgGlnProProAsp 140
QY	3861	CGATGGATGACCAAGCCCGCATGACCCACTATCAAGCCTGCTTCTCACAGAGGGTTC 3920
Db	141	ArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuAsnGluArgVal 160
QY	3921	ACGTTGCTGCCAGCGCGTCTCAACCTGCCACTCTTCGCTGGAAGACTGATGAA 3980
Db	161	SerpPheAlaProAlaValLeuAsnProAlaThrLeuLeuProValGluSerGlnAla 180
QY	3981	CCAGTGACTCATGATGTCATCACTATGATTGAGGAGACTGGGTCGCGCAAGACCTT 4040
Db	181	ThrProValHisArgCysSerGluIleLeuAlaGluGluThrGlyThrArgArgAspLeu 200
QY	4041	ACAGACATACCGCTGACTGGAGAAGTCTAAACCTGCTTCACTGACGGAAGCAGCTATGTG 4100
Db	201	LysAspGlnProLeuProGly---ValProAlaTrpTyrThrAspGlySerSerPheIle 219
QY	4101	GTGAAGGTAAAGAGATGGTGGGCGCGTGTGTGACGGGACCCGACCATCTGGGCC 4160
Db	220	AlaGluGlyLysArgAlaGlyAlaAlaIleValAspGlyLysArgThrValTrpAla 239
QY	4161	AGCAGCTTCGCGGAGGAACTTCAGCAAAAGCTGAGCTCATGGCCCTCACCCAGCT 4220
Db	240	SerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuValAlaLeuThrGlnAla 259
QY	4221	TTGGCGCTGGCGGAAGGAAATCCATAAATTTATACGGACAGCAGGTATGCTCTTTCGC 4280
Db	260	LeuArgLeuAlaGluGlyArgAspIleAsnIleTyrThrAspSerArgTyrAlaPheAla 279
QY	4281	ACTGCACAGTACATGGGGCCATCTATAAACAAGGGGTGCTTTACCTCAGCAGGGAGG 4340
Db	280	ThrAlaHisIleHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyLys 299
QY	4341	GAATAAGAACAAAGAGGAATCTAAGCTTATTAGACCGGTACATTTACCAAAAGG 4400
Db	300	AspIleLysAsnLysGluGluLeuAlaLeuLeuGluAlaIleHisLeuProLysArg 319
QY	4401	CTAGCTATTATACACTGTCCTGGACATCAGAAAAGCTAAAGATCTCATATCCAGAGGAAC 4460

Db	320	ValAlaIleIleHisCysProGlyHisGlnLysGlyAsnAspProValAlaThrGlyAsn 339
QY	4461	CAGATGGCTACCGGGTTCGACAGCAGCGCCAGGGTGTAACTCTTCGCTATATA 4520
Db	340	ArgArgAlaAspGluAlaAlaLysGlnAlaLeuSerThrArgValLeuAlaGluThr 359
QY	4521	GAATGCCCCAAAGCCCGAGAACCCAGACGACATCACCTAGAGAAGCTGGCAAGAGATA 4580
Db	360	ThrLysPro-----GlnGluLeu 365
QY	4581	AAAAAGATAGACCACTTCTCTGAGACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAG 4640
Db	366	IleLysProAlaGlnValLysThrArgPro-----Gly 376
QY	4641	GAATCTGCCCCCAAGAAGGGTTAGATATGTCACACAGATACATCGTTACCCAC 4700
Db	377	GluLeuThrProAspArg---GlyLysGluPheIleGlnArgLeuHisGlnLeuThrHis 395
QY	4701	CTAGGAATAAACACCTTCGTCAGACAGTTCGTGAGACATCCCCCTTATCATGTTCTCAGGCTA 4760
Db	396	LeuGlyProGluLysLeuLeuGlnLeuValAsnArgThrSerLeuLeuIleProAsnLeu 415
QY	4761	CCAGAGTGGCTGACTCGGTGTCAAACATTTGTGTGTCGCCAGCTGGTTAATGCTAAT 4820
Db	416	GlnSerAlaValArgGluValThrSerGlnCysGlnAlaCysAlaMetThrAsnAlaVal 435
QY	4821	CCRTCCAGATGCTCCAGGGAGAGACTAAGGGAGAGCCACCCAGCGCTCACTGGGAA 4880
Db	436	ThrThrTyrArgGluThrGlyLysArgGlnArgGlyAspArgProGlyValTyrTrpGlu 455
QY	4881	GTGACATCTCACTGAGGTAAAGCCGCTAAATACCGAAACAAATACCTATTGTTTGTGA 4940
Db	456	ValAspPheThrGluValLysProGlyArgTyrGlyAsnArgTyrLeuLeuValPheIle 475
QY	4941	GAC 4943
Db	476	Asp 476
RESULT 13		
A42743		
pol polyprotein - radiation murine leukemia virus (strain Kaplan) (fragment)		
N;Contains: endonuclease (EC 3.1.1.-); proteinase (EC 3.4.21.-); RNA-directed DNA polym		
C;Species: radiation murine leukemia virus		
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004		
C;Accession: A42743		
R;Poliquin, L.; Bergeron, D.; Fortier, J.L.; Paquette, Y.; Bergeron, R.; Rassart, E.		
J. Virol. 66, 5141-5146, 1992		
A;Title: Determinants of thymotropism in Kaplan radiation leukemia virus and nucleotide		
A;Reference number: A42743; MUID:92333703; PMID:1629969		
A;Accession: A42743		
A;Molecule type: DNA		
A;Residues: 1-581 <POL>		
A;Cross-references: UNIPROT:P31795; UNIPARC:UPI0000131P09; GB:M93052; NID:g332065; PDB:		
C;Comment: This protein is likely to be expressed as a fused gag-pol polyprotein.		
C;Comment: The precise boundary between the mature proteins has not been determined.		
C;Genetics:		
A;Gene: pol		
C;Superfamily: pol polyprotein		
C;Keywords: endonuclease; hydrolase; nucleotidyltransferase; polyprotein; reverse trans		
Alignment Scores:		
Pred. No.:	1.73e-95	Length: 581
Score:	1681.00	Matches: 329
Percent Similarity:	69.9%	Conservative: 84
Best Local Similarity:	55.7%	Mismatches: 148
Query Match:	11.5%	Indels: 30
DB:	2	Gaps: 6
US-10-723-552-3 (1-8132) x A42743 (1-581)		
QY	4002	CAACTATTGATGTGAGGAGACTGGGTCCGGAAGACCTTACAGACATACCTGACTGGA 4061
Db	2	GluiteLeuAlaGluThrHisGlyThrArgProAspLeuThrAspGlnProIleProAsp 21





Query Match:	11.2%	Indels:	280
DB:	2	Gaps:	42
US-10-723-552-3 (1-8132) x T09394 (1-1751)			
QY	939	AAACACTCGCGTGAAGAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGACCA	998
DB	177	LySArgAspGluLySLeuLeuProAspThrThrGlnLySValGlnThr	196
QY	999	CCGCTTGGCGGAAACCAATCTGTTCCCCACCCCTTATCTGGCACAGGTTGCCGCG	1058
DB	197	AspIleGlyGluAlaSerSer	208
QY	1059	AGGGACCTTTGCCCCCTCTGGAGCTCCGCGGTGGAGGACCTGCTGAGGACCTCGG	1118
DB	209	GlnLySProIleSerThrAspAsnProAspLeuLySValAspGlyValLeuThrArg	228
QY	1119	AGCCGGAGGGCGCCACCCCG-----GAGCGGACAGCAGATC	1157
DB	229	SerGlnHisThrThrValProSerAsnIleThrIleLySAspGlyThrSerValGln	248
QY	1158	GGCAGATTACGCTGCGCAGCTACGCGCTCCACACCGGGGCCAATTTGCAGCCCTC	1217
DB	249	TyrGlnHisProIleArgAsnTyr-----ProThrGlyGluGlyAsnLeuThrAlaGln	266
QY	1218	CAGTATTGGCCCTTTCTCTGCAGATCTCTATAATTGAAACTAACCATCCCTCTTC	1277
DB	267	ValArgAsnProPheArgProLeuGluLeuGlnLeuArgLySAspCysProAlaLeu	286
QY	1278	TCGGAGGATCCCAACCCCTCACGGGTGTGGAGTCCCTTATGTTCTCTCACCAGCCT	1337
DB	287	ProGluGlyIleProGlnLeuAlaGluTrpLeuThrGlnThrMetAlaIleTyrAsnCys	306
QY	1338	ACTTGGGATGATTCTCAACAGCTCTGCAGACACTCTTCAACCGGAGGACGAGAGA	1397
DB	307	AspGluAlaAspValGluGlnLeuAlaArgValIlePheProThrProValArgGln---	325
QY	1398	ATTCTATTAGAGGCTAGAAAAAATGTTCTTGGGGCGGACCGGCGCCACCGCGGTGCA	1457
DB	326	-----IleAlaGlyValIleAsnGlyHisAlaAlaAlaAsnThrAlaAlaLySleGln	343
QY	1458	AATGAGATT-----GACATGGGATTTCCCTTA	1484
DB	344	AsnTyrValThrAlaCysArgGlnHisTyrProAlaValCysAspTrpGly-----	360
QY	1485	ACTCGCCCCGTTGGGACTACACACGCTGAAGTAGGAGGACCTTGAAAACTATCGC	1544
DB	361	ThrIleGlnAlaPheThrTyrLySProProGlnThrAlaHisGluTyrValLySHisAla	380
QY	1545	CAGGCTCTGTGGGGGTCTCCGGGGCGCCTCAAGACGGCCCACTAATTTGGCTAAGTA	1604
DB	381	GluIleIlePheLySAsnAsnSerGlyLeuGluTrpGlnHisAlaThrValProPheIle	400
QY	1605	AGAGAGTGTAGGAGGACCGAATGAACCCCTCTGTT-----	1643
DB	401	AsnMetValValGlnGly-----LeuProLySValThrArgSerLeuMetSerGly	418
QY	1644	-----TTTCTTGAGAGGCTCTTGAAGCCTTCAGCGGTACACC	1682
DB	419	AsnProAspTrpSerThrLySThrIleProGlnIleIleProLeuMetGlnHisTyrLeu	438
QY	1683	CCTTTTGATCCCACTCAGAGGCCCAAAAAAGCCCTCAGTGGCTTTTGGCCCTTATAGGACAG	1742
DB	439	AsnLeuGlnSerArgGlnAspAlaLySleIleLySlnThrProLeu-----	453
QY	1743	TCAGCTTGATATTAGAAAGAGCTTCAGAGACTGGAAGGGTTACAGGAGGCTGAGTTA	1802
DB	454	---ValLeuGlnLeuAlaMetProAlaGlnThrMetAsnGly-----	466
QY	1803	CGTGATCTAGTGAAGGCGCAGAGAAGTATATTACAAAGGGGACAGAGAAGAAAGG	1862
DB	467	-----AsnLySgLyTyrValGlySerTyrProThrAsnGluPro	479
QY	1863	GAACAAAGAAAG	1922
DB	480	TyTyrSerPheGlnGlnGlnGln-----	487
QY	1923	AAGAAATTTGACTAAGATCTTGGCTGCGAGTGGTTGAAGGGAAGCAATACGAAAGAGAG	1982
DB	487	-----	487
QY	1983	AGAGATTTTAGAAAAATTAGTCTAGGCCCTAGA---CAGTCAGGGAACCTGGGCAATAGG	2039
DB	488	-----ArgProAlaProArgAlaProProGlyAsnValProSerAsn	501
QY	2040	ACCCCACTGCACAAAGACCAATGTGCATATTGTAAAGAAAGAGACACTGGGCAAGAAC	2099
DB	502	Thr-----CysPheCysLySlnProGlyHisTyrLySAlaAsp	515
QY	2100	TGCCCCNAGAG-----GGAAACAAGGA	2123
DB	516	CysProAsnLySThrArgAsnLeuArgAsnMetGlyAsnMetGlyArgGlyArgMet	535
QY	2123	-----	2123
DB	536	GlyGlyProProTyrArgSerGlnProTyrProAlaPheIleGlnProProGlnAsnHis	555
QY	2124	-----CCAAGGATCCTAGCTCTAGAGAGATATAA	2153
DB	556	GlnAsnGlnTyrAsnGlyArgMetAspArgSerGlnLeuGlnAlaSerAlaGlnGluTrp	575
QY	2154	GATTAGGGAGCGGGTTCGGACCCCTCCCGAGGCC-----	2192
DB	576	LeuProGlyThrTyrProAlaAspProIleAspCysProTyrGluLySArgGlyThrLyS	595
QY	2193	-----AGGTAACCTTTGAAGTGGAGGCGGCA	2219
DB	596	ThrThrGlnAspValIleThrThrLySAsnAlaGluIleMetValThrValAsnHisThr	615
QY	2220	CCAGTTGAGTTCCTGTTGATACCGGAGCGAAACATTCAGTGCTA-----	2264
DB	616	LySleProMetLeuValAspThrGlyAlaCysLeuThrAlaIleGlyGlyAlaAlaThr	635
QY	2265	CTACAGCCATTAGGAAAACCTAAAGATATAAAATCTCGGTGATGGTGCACAGGGCAA	2324
DB	636	ValValProAspLeuLySLeuThrAsnThrGluIleIleAlaValGlyIleSerAlaGlu	655
QY	2325	CAACAGTATCCATGGACTACCCGGAAGACA-----GTTGACTTGGAGTGGGACGGTAACC	2381
DB	656	ProValProHisValLeuAlaLySProThrLySleIleGlnIleGluAsnThrAsnIleAsp	675
QY	2382	CACCTCGTTTCTGCTCATACCTGAGTGCACGACCCCTCTTAGGTAGAGACTTATTGACC	2441
DB	676	IleSerProTyrTyrAsnProAspGlnThrPheHisIleLeuGlyArgAspThrLeuSer	695
QY	2442	AGATGGGAGACAAATTTCTTTTGAACAAGGAAACCGAAGTGTCTGCAATAACAAA	2501
DB	696	LySMetArgAlaIleValSerPheGluLyS-----AsnGly	707
QY	2502	CCTATCAGCTGTGTGACCCCTCCAATTAGATGACGAATAT---CGACTATACTCTCCCTA	2558
DB	708	GluMetThrValLeu-----LeuProProThrTyrHisLySlnLeuSerCysGln	724
QY	2559	GTAAGCCTGATCAAAATATACAAATCTGTTGGTGAACAGTTTCCC---CAAGCCTGGGCA	2615
DB	725	ThrLySAsnThrLeuAsnIleAspGluTyrLeuLeuGlnPheProAspGlnLeuTrpAla	744
QY	2616	GAACCGCAGGATGGTTGGCAAGCAAGTTCCTCCCAACAGTATTTCACCTGAAGGCC	2675
DB	745	SerLeuProThrAspIleGlyArgMetLeuValProIleThrIleLySleIleLySAsp	764
QY	2676	AGTGCCACACCATGCTCAGTCAGACAGATCCCTTCAGTAAAGAGCTCAAGAGGAATT	2735
DB	765	AsnAlaSerLeuProSerIleArgGlnTyrProLeuProLySAspLySThrGluGlyLeu	784
QY	2736	CGGCCGCATGTCCAAAGATTAATCCACAGGGCATCTAGTTCTCTCTCCCAATCTCCCTGG	2795



Db 785 ArgProLeuIleSerSerLeuGluAsnGlnGlyLeuLeuLysCysHisSerProCys 804  
Qy 2796 AATACTCCCTGCTACCGTTAGAAAGCTGGGACTAATGACTATTCAGACCGATACAGAC 2855  
Db 805 AsnThrProIlePheProIleLysLeuAlaGlyArgAspGluTyrArgMetIleHisAsp 824  
Qy 2856 TTGAGAGGTCATAAACCAGGTGCAGATATACACCAACAGTCCCGAACCCCTTATAAC 2915  
Db 825 LeuArgAlaIleAsnAsnIleValAlaProLeuThrAlaValAlaValAlaSerProThrThr 844  
Qy 2916 CTCTGTGTGCTCTCCACCCCAACGAGCTGGTATACAGTATTGGACTTTAAAGGATGCC 2975  
Db 845 ValLeuSerAsnLeuAlaProSerLeuHisTrpPheThrValIleAspLeuSerAsnAla 864  
Qy 2976 TTCTTCCTGCCTGAGATTACACCCCACTAGCCACCACTTTTGGCTTCGATGGAGAT 3035  
Db 865 PhePheSerValProIleHisLysAspSerGlnTyrLeuPheAlaPhe----- 880  
Qy 3036 CCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCGACTGCCCAAGGTTTCAAGAAC 3095  
Db 881 -----ThrPheGluGlyHisGlnTyrThrTrpValLeuProGlnGlyPheIleHis 898  
Qy 3096 TCCCGCAGCACTTTTGACAGCCCTACACAGACACTG-----GCCAATCTCAGGATC 3149  
Db 899 SerProThrLeuPheSerGlnAlaLeuTyrGlnSerLeuHisLysIleLysPheLysIle 918  
Qy 3150 CAACACCTCAGGTAGCCCTCCTCCAGTACGTGATGACCTGCTTCTGGCGGAGCCACC 3209  
Db 919 SerSer-----GluIleCysIleTyrMetAspValLeuIleAlaSerLysAsp 935  
Qy 3210 AACAGGACTCTGTAGAGGACGACGACTACTGTGGAATGTCTGACCTAGGCTAC 3269  
Db 936 ArgAspThrAsnLeuLysAspThrAlaValMetLeuGlnHisLeuAlaSerGluGlyHis 955  
Qy 3270 AGACCTCTGTGAAGAAGCCAGATTTCAGGAGAGAGGTAAACATACTTTGGGTACAGT 3329  
Db 956 LysValSerLysLysLysLeuGlnLysCysGlnGlnGluValTyrLeuGlyGlnLeu 975  
Qy 3330 TTGGCGGACGGCAGCGATGCTGACGAGGACGCGAAGAAACTGTAGTCCAGATACCG 3389  
Db 976 LeuThrProGluGlyArgLysIleLeuProAspArgLysValThrValSerGlnPheGln 995  
Qy 3390 GCCCCAACCCACGACCAAACTAGCAGAGTTCCTGGGACAGCTGATTTTGGAGACTG 3449  
Db 996 GlnProThrThrIleArgGlnIleArgAlaPheLeuGlyLeuValGlyTyrCysArgHis 1015  
Qy 3450 TGGATCCCGGGTTTGGACCTTAGCAGCCCACTC---TACCCGCTAACCAAGAAAA 3506  
Db 1016 TrpIleProGluPheSerIleHisSerLysPheLeuGluLysGlnLeuLysLysAspThr 1035  
Qy 3507 GGGGAA---TTCCTCTGGGCTCTGAGCAGCAGAGGATTTGATGCTATCAAAAAGGCC 3563  
Db 1036 AlaGluProPheGlnLeuAspAspGlnGlnValGluAlaPheAsnLysLeuLysHisAla 1055  
Qy 3564 CTGCTGACGACCTGCTGCGCCCTCCCTGACCTAACTAAACCTTTTACCTTTATGTG 3623  
Db 1056 IleThrThrAlaProValLeuValProAspProAlaLysProPheGlnLeuTyrThr 1075  
Qy 3624 GATGAGCTTAAGGAGTAGCCGGGGAGTTTAAACCCCAACCCCTAGGACCATGAGAGA 3683  
Db 1076 SerHisSerGluHisAlaSerIleAlaValLeuThrGlnLysHisAlaGlyArgThrArg 1095  
Qy 3684 CCTGCTGCTACCTGTCAAGAAGCTGATCCTGTAGCAGTGGTGGCCCATATGCTG 3743  
Db 1096 ProIleAlaPheLeuSerSerLysPheAspAlaIleGluSerGlyLeuProProCysLeu 1115  
Qy 3744 AAGGCTATCGAGCTGTGGCCATACTGTCAGGACCTGCAAAATTCACCTTTGGGACAG 3803  
Db 1116 LysAlaCysAlaSerIleHisArgSerLeuThrGlnAlaAspSerPheIleLeuGlyAla 1135  
Qy 3804 AATATACTGTAATAGCCCCCATGTCATTGGAGAACATCGTTTCGGCAGCCCCCAGACCGA 3863  
Db 1135

Db 1136 ProLeuIleIleTyrThrThrHisAlaIleCysThrLeuLeuGlnArgAspArgSerGln 1155  
Qy 3864 TGGATGACCAACGCCCGCATGACCCACTATCAAAAGCTGCTCTCTCACAGAGAGGTCACG 3923  
Db 1156 LeuValThrAlaSerArgPheSerLysTrpGluAlaAspLeuLeuArgProGluLeuThr 1175  
Qy 3924 TTGCTCCACAGCGCTCTCAACCCCTGCCACTCTCTGCTCGAAGAGACTGTAGAACCA 3983  
Db 1176 PheValAlaCysSerAlaValSerProAlaHisLeuTyrMetGlnSerCysGluAsnAsn 1195  
Qy 3984 GTG---ACTCATGATGCCATCAACTATTGATTGAGGAGACTGGGGTCGCAAGACCTT 4040  
Db 1196 IleProProHisAspCysValLeuLeuThrHisThrIleSerArgProArgProAspLeu 1215  
Qy 4041 ACAGACATACCGCTGACTGGAGAAGTCTAACTGTTCTACTGACGGAACGACTATGTG 4100  
Db 1216 SerAspLeuProIleProAspProAspMetThrLeuPheSerAspGlySerTyrThrThr 1235  
Qy 4101 GTGGAAGGTAAAGAGATGGCTGGCGCGCTGTGACGGGACCCCGCACG----- 4151  
Db 1236 GlyArgGly-----GlyAlaAlaValValMetHisArgProValThrAspAsp 1251  
Qy 4152 ---ATCTGGCCAGACGCTGCCGGAAGAACTTTCAGCAACAAAGGCTGAGCTCATGGCC 4208  
Db 1252 PheIleIleIleHisGlnGlnProGlyGlyAlaSerAlaGlnThrAlaGluLeuLeuAla 1271  
Qy 4209 CTCACGCAAGCTTTGCGCTGGCGGAGGAAATCCATAACATTTATACGACGACGACGAG 4268  
Db 1272 LeuAlaAlaCysHisLysAlaThrAspLysThrValAsnIleTyrThrAspSerArg 1291  
Qy 4269 TATGCTTTTCGACTGCACACTGACATGCGGCGCATCTATAACAAAGGGGTTCTTACC 4328  
Db 1292 TyrAlaTyrGlyValValHisAspPheGlyHisLeuTrpMetHisArgGlyPheValThr 1311  
Qy 4329 TCACGAGGGAGGAAATAAGAACAAAGAGAAATTTAAAGCTTATTAGAACCGTCATCAT 4388  
Db 1312 SerAlaGlyThrProIleLysAsnHisLysGluIleGluTyrLeuLeuLysGlnIleMet 1331  
Qy 4389 TTACCAAAAGGCTAGCTATTATACACTGCTCGACATCAGAAAGCTTAAAGATCTCATA 4448  
Db 1332 LysProLysGlnValSerValIleLysIleGluAlaHisThrLysGlyValSerMetGlu 1351  
Qy 4449 TCACAGAGAAACACAGATGGCTGACCGGTTGCCAAGCAGCAGCCCGAGGTGTTAACCTT 4508  
Db 1352 ValArgGlyAsnAlaAlaAlaAspGluAlaAlaLysAsnAlaValPheLeuVal----- 1369  
Qy 4509 CTGCTTAATATAGAAATGCCAAAGCCCAAGCCAGACCCAGACAGACAGTACCTAGAGAC 4568  
Db 1370 -----GlnArgValLeuLysLys 1375  
Qy 4569 TGGCAAGAGATAAAAAGATAGAC-----CAGTTCTCTGAGACTCCGGAAGGGAC 4619  
Db 1376 GlyAspAlaLeuAlaSerThrAspLeuValMetGluTyrSerGluThrAspGluLysPhe 1395  
Qy 4620 TGCTATACCTCAGAT-----GGGAAGGAATCTCTCCCCAC 4655  
Db 1396 ThrAlaGlyAlaGluLeuHisAspGlyValPheMetArgGlyAspLeuIleValProPro 1415  
Qy 4656 AAAGAAGGTTAGAAATATGTC---CAACAGATACATCGTCTAACCCCACTA----- 4703  
Db 1416 LeuGluMetLeuHisAlaIleLeuLeuAlaIleHisGlyValSerHisThrHisLysGly 1435  
Qy 4704 GGAATAAACACCTGACAGTGGTTCAGAACATCCCTTATCATGTTCTGAGGCTACCA 4763  
Db 1436 GlyIleMetSerTyrPheSerLysPheThrHisPro-----LysAlaSer 1451  
Qy 4764 GGAGTGGCTACCTCGGTGGTCAAACTGTGTGCTCCAGCTGGTGGTAAATCTTAACTCT 4823  
Db 1452 GlnThrIleAspLeuIleLeuGlyHisCysGlnIleCysLeuLysHisAsnProLysTyr 1471  
Qy 4824 TCCAGATGCTCCAGGAAGAGACTAAGGGAGAACCCAGCCAGGCGCTCATCTGGGAAGTG 4883  
Db 1472 LysSerArgLeuGlnGlyHisArgProLeuProSerArgProPheAlaHisLeuGlnIle 1491



```
Qy 4884 GACTTCACTGAGTA-----AAGCCGGCTAAATACGGAACAATACTATTGGTT 4934
Db      ||||| :|||:
1492 AspPheValGlnMetCysValLysPro-----MetTyrAlaLeuVal 1506
Qy 4935 TTTCAGACACCTTTTCAGGATGGGTAGAGCCTTATCCTACTAAGAAGAGACTTCAACC 4994
Db      ||||| :|||:
1507 IleIleAspValPheSerLysTrpProGluIleProCysAsnLysGluAspAlaLys 1526
Qy 4995 GTGGTGGCTAAATAAATACTCGAAGAAATTTTCCAGATTTCGAATACCTAAGGTAATA 5054
Db      ||||| :|||:
1527 ThrValCysAspIleLeuMetLysAspIleIleProArgTrpGlyLeuProAspGlnIle 1546
Qy 5055 GGGTCAGACAAATGTCAGCTTTTGGTCCAGCTAGTACAGGACTGGCCCAAGATATTG 5114
Db      ||||| :|||:
1547 AspSerAspGlnGlyThrHisPheThrAlaLysIleSerGlnGluLeuThrHisSerIle 1566
Qy 5115 GGGATTGATGGAACCTGCATTGTGCATACAGACCCCAAGCTCAGACAGGTAGAGAGG 5174
Db      ||||| :|||:
1567 GlyValAlaTrpLysLeuHisCysProGlyHisProArgSerSerGlyIleValGluArg 1586
Qy 5175 ATGAATAGAACCAATAAAGAGACCTTACTAAATTGACCGCGGAGACTGGCGTTAATGAT 5234
Db      ||||| :|||:
1587 ThrAsnArgThrLeuLysSerLysIleIleLysAlaGlnGluGlnLeuLysSerLys 1606
Qy 5235 TGGATAGCTCTCCCTGCTTTGTGCTTTTGGGTTAGGTTAGGAACACCCCTGGACAGTTTGGG 5294
Db      ||||| :|||:
1607 TrpThrGluValLeuProTyrValLeuLeuGluMetArgAlaThrProLysLysHisGly 1626
Qy 5295 CTGACCCCTTATGATTAATCTACTACGGGGGACCCCCCA-----TTGGTAGAAATT 5345
Db      ||||| :|||:
1627 LeuSerProHisGluIleValMetGlyArgProMetLysThrThrTyrLeuSerAspMet 1646
Qy 5346 GCTTCTGTACATAGTGTGCTGCTTTCCAGCCTTTGTTCTCTAGGCTCAAGGCA 5405
Db      ||||| :|||:
1647 SerProLeuTrpAlaThrAspThrLeuValThr----- 1657
Qy 5406 CTTGAGTGGGTGAGACAAACGAGCGTGGAGGCACTC-----CGGAGGCGCTAC 5453
Db      ||||| :|||:
1658 -----TyrMetAsnLysLeuThrArgGlnLeuSerAlaTyrHisGlnGlnValVal 1674
Qy 5454 TCAGGAGGAGGACTTGCAGATCCACATCGTTTCCAAAGTGGGAGATTTCAGTCTAGTT 5513
Db      ||||| :|||:
1675 AspGlnTrpProSerThrSerLeuProProGlyProGluProGlySerTrpCysMetLeu 1694
Qy 5514 AGACGCCACCGTGCAGGAACCTCGAGACTCGGTGGGAAGGCCCTTATCTCGTACTTTTG 5573
Db      ||||| :|||:
1695 ArgAsnProLysLys-----SerSerAsnTrpGluGlyProPheLeuIleLeuLeu 1711
Qy 5574 ACCACACCAACGGCTGTGAAGTCGAAGGAATCTCCACCTGGATCCATGCCACGTT 5633
Db      ||||| :|||:
1712 SerThrProThrAlaValLysValGluGlyArgProThrTrpIleHisLeuAspHisCys 1731
Qy 5634 AAA 5636
Db      ||||
1732 Lys 1732
```

Search completed: February 14, 2006, 16:02:52  
Job time : 502.429 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 12:52:14 ; Search time 338.343 Seconds  
(without alignments)  
3391.447 Million cell updates/sec

Title: US-10-723-552-3

Perfect score: 14636

Sequence: 1 GCGTGGGTGACTGACTGGG.....CTGTTTCATCAAAAAA 8132

Scoring table:

	BLOSUM62
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB spool/US10723552/runat 14022006 125141 12876/app query.fasta\_1  
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p  
-USR=US10723552 @CGN 1.1 808 -runat 14022006 125141 12876 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONLOG -DEV TIMEOUT=120  
-WEN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -XGAPOP=6 -XGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11454.5	78.3	2378	2 Q91X4_9GAMR	Q91X4 porcine end
2	11005.5	75.2	2376	2 Q91X5_9GAMR	Q91X5 porcine end
3	10997.5	75.1	2376	2 Q91X3_9GAMR	Q91X3 porcine end
4	8821	60.3	1720	2 Q8J4V6_9GAMR	Q8J4V6 porcine end
5	8815.5	60.2	1718	2 Q8J4V8_9GAMR	Q8J4V8 porcine end
6	6248.5	42.7	1193	2 Q5QG8_9GAMR	Q5QG8 porcine end
7	6170.5	42.2	1193	2 Q90RL9_9GAMR	Q90RL9 porcine end
8	6109	41.7	1195	2 Q90RL6_9GAMR	Q90RL6 porcine end
9	6026	41.2	1145	2 Q9XSN8_FIG	Q9XSN8 sus scrofa
10	5986	40.9	1145	2 Q4VFZ2_9GAMR	Q4VFZ2 porcine end
11	5863	40.1	1146	2 Q8UMF5_9GAMR	Q8UMF5 porcine end
12	5840.5	39.9	1144	2 Q8UM99_9GAMR	Q8UM99 porcine end
13	5831	39.8	1147	2 Q8UM96_9GAMR	Q8UM96 porcine end
14	5466.5	37.3	1142	2 Q8Q6U4_9GAMR	Q8Q6U4 porcine end
15	5383.5	36.8	1139	2 Q8Q6U7_9GAMR	Q8Q6U7 porcine end
16	5383.5	36.8	1736	2 Q7ZJT6_9GAMR	Q7ZJT6 amphotrophic

17	5383	36.8	1733	2 Q9E7M1_9GAMR	Q9E7M1 dg-75 murin
18	5347.5	36.5	1734	2 Q9J8E2_9GAMR	Q9J8E2 murine leuk
19	5344.5	36.5	1738	2 Q39735_MLVFR	Q39735 friend muri
20	5343.5	36.5	1734	2 Q7SVK7_9GAMR	Q7SVK7 murine leuk
21	5338	36.5	1737	2 Q92808_MLVMO	Q92808 moloney mur
22	5333.5	36.4	1738	2 Q8UN00_MLVMO	Q8UN00 moloney mur
23	5333.5	36.4	1734	2 P70355_MOUSE	P70355 mus musculu
24	5291.5	36.2	1736	2 Q5PYI3_9GAMR	Q5PYI3 murine leuk
25	5266.5	36.0	1736	2 Q83362_9GAMR	Q83362 murine leuk
26	5255	35.9	1786	2 Q89811_FLV	Q89811 feline leuk
27	5237	35.8	1784	2 Q85521_FLV	Q85521 feline leuk
28	4450	30.4	868	2 Q73505_9GAMR	Q73505 porcine end
29	4332.5	29.6	1203	2 Q89815_9GAMR	Q89815 mus dunn e
30	4316.5	29.5	1165	1 POL_GALV	P21414 gibbon ape
31	4177.5	28.5	1127	2 Q9TTC1_PHACI	Q9TTC1 phascolatct
32	4154.5	28.4	1127	2 Q70652_GALV	Q70652 gibbon ape
33	3996	27.3	1196	1 POL_MLVAV	P03356 akv murine
34	3992.5	27.3	1204	1 POL_MLVFP	P26808 friend muri
35	3987.5	27.2	1199	2 Q60FS9_9GAMR	Q60FS9 murine leuk
36	3986.5	27.2	1199	2 Q60FS6_9GAMR	Q60FS6 murine leuk
37	3984.5	27.2	1204	2 Q41250_MLVRA	Q41250 rauscher mu
38	3981	27.2	1196	2 Q90RL4_9GAMR	Q90RL4 murine leuk
39	3980.5	27.2	1204	1 POL_MLVFF	P26809 friend muri
40	3977.5	27.2	1204	2 Q9YK99_9GAMR	Q9YK99 murine leuk
41	3969.5	27.1	1199	1 POL_MLVMO	P03355 moloney mur
42	3959.5	27.1	1204	1 POL_MLVF5	P26810 friend muri
43	3947	27.0	1196	1 POL_MLVRD	P11227 radiation m
44	3940.5	26.9	1189	1 POL_BAEVM	P10272 baboon endo
45	3939	26.9	1204	2 Q7ZKZ7_9GAMR	Q7ZKZ7 recombinant

#### ALIGNMENTS

RESULT 1  
Q91X4\_9GAMR  
ID Q91X4\_9GAMR PRELIMINARY; PRT; 2378 AA.  
AC Q91X4;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Type C proviral gag, pol and env genesand LTR (Class A, clone 42).  
OS Porcine endogenous retrovirus.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gammaetrovirus; 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN-type C;  
RA Czauderna F., Fischer N., Boller K., Krach U., Kurth R., Toenjes R.R.;  
RT "Molecular Characterization of Human-tropic and Replication-competent  
RT Porcine Endogenous Retroviruses.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ13817; CAB65340.1; -; Genomic\_DNA.  
DR HSSP; P03385; 1MOF.  
DR SMR; Q91X4; 7-98, 2227-2279.  
DR MEROPS; A02\_020; -;  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006310; F:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0019068; P:RNA assembly; IEA.  
DR InterPro; IPR002050; Env polyprotein.  
DR InterPro; IPR008981; FmLVrecept-bind.  
DR InterPro; IPR003036; Gag\_p30.  
DR InterPro; IPR000840; G retro matrix.  
DR InterPro; IPR001995; Peptidase\_A2\_cat.

```

DR InterPro; IPR000223; Peptidase_S26A.
DR InterPro; IPR001969; Pept_Asp_AS.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01140; Gag_MA; 1.
DR Pfam; PF02093; Gag_P30; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT_1; 1.
DR Pfam; PF00429; TLV_coat; 1.
DR Pfam; PF00098; zf_CCHC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR PROSITE; PS50879; RNase H; 1.
DR PROSITE; PS00501; SPASE_1; 1.
DR PROSITE; PS50158; ZF_CCHC; 1.
DR CHAIN 1 524 Gag protein.
FT CHAIN 525 1759 Pol protein.
FT CHAIN 1719 2378 Env protein.
SQ SEQUENCE 2378 AA; 267306 MW; 6EC8DAEC1B4ED18A CRC64;

Alignment Scores:
Pred. No.: 0 Length: 2378
Score: 11454.50 Matches: 2156
Percent Similarity: 94.1% Conservative: 73
Best Local Similarity: 91.0% Mismatches: 77
Query Match: 78.3% Indels: 64
DB: 2 Gaps: 8

US-10-723-552-3 (1-8132) x O9Q1X4_9GAMR (1-2378)
QY 585 ATGGGACAGCGGTGACGACCCCTCTTAGTTGACTCTCGACCATTTGACATGAAGTTAAA 644
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20
QY 645 TCCAGGGCTCATAATTTTCAGTTTCAAGTAAAGGACCTTGGCAGACTTTCTGTGTC 704
Db 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysAla 40
QY 705 TCTGAATGGCGACATTCGATGTGGATGGCCATCAGAGGGACCTTTAATCTCGAGATT 764
Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60
QY 765 ATCTGGCTGTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 824
Db 61 IleLeuAlaValLysAlaIlelePheGlnThrGlyProGlySerHisProAspGlnGlu 80
QY 825 CCCTATATCTTACGTGGCAAGATTGGCAGAGGATCTCCGCCATGGGTTAAACCATGG 884
Db 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100
QY 885 CTGAATAAGCCAAAGAACCGAGTCCCGAATCTGGCTCTTGGAGAGAAAACAACAC 944
Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120
QY 945 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGAGCCACCGGCT 1004
Db 121 SerAlaGluLysValGluProSerProArgIleTyrProGluIleGluGluProProThr 140
QY 1005 TGGCGGGAACCCCAATCTGTTCCCCACCCCTTATCTGGCACAGGGTGGCGGAGGGA 1064
Db 141 TrpProGluProGlnProValProProProTyrProAlaGlnGlyAlaValArgGly 160
QY 1065 CCTTTTGGCCCTCTGGAGCTCCGGCGGTGAGGGACCTGCTGCGGGACTCGAGCGCG 1124
Db 161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180
QY 1125 AGGGGCGCCACCCGGAGCGGACGACGAGATCCGACATTTACCGCTCGCGCACGTACGGC 1184

```

```

Db 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly 200
QY 1185 CTTCCCAACACGGGGGCCAATTGTCAGCCCTCAGATTTCGCCCTTTTCTTCGAGAT 1244
Db 201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220
QY 1245 CTCTATAATTGGAAAACTAACCATCCCTTTCGCGAGGATCCCAACGCGCTCACGGGG 1304
Db 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240
QY 1305 TTGCTGAGTCCCTTATGTTCTCTCACAGCTACTTTGGGATGATTGTCAACAGCTGCTG 1364
Db 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeu 260
QY 1365 CAGACATCTTCAACACCGAGGAGGAGAGAAATTCATTATAGAGGCTAGAAAAATGTT 1424
Db 261 GlnThrLeuPheThrThrGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280
QY 1425 CTTGGGGCCGACGGCGCACCCAGCGGTTCGAAATCGAGATTGACATGGGATTTCCCTTA 1484
Db 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300
QY 1485 ACTGCGCCCGGTTGGGACTACACACGCGCTCAAGGTAGGAGAGCTTCAAAATCTATCGC 1544
Db 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320
QY 1545 CAGCTCTGTGGCGGTCTCCGGGGCGCTCAAGACGGCCCTCAATTTGGCTAAGGTA 1604
Db 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340
QY 1605 AGAGAAGTGTATGACGAGGACCGAATGAACCCCTCTCTTTTCTTTGAGAGGCTCTTGAA 1664
Db 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
QY 1665 GCCTTCAGGGCGGTACACCCCTTTTGATCCACCTCAGAGGCGCCAAAAGCCTCAGTGGCT 1724
Db 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380
QY 1725 TTGGCCCTTTATAGACAGTCCAGCCCTTGGATATTAGAAGAGCTTCAGAGCTCGAAGGG 1784
Db 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400
QY 1785 TTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAGCGCAGAGAAGTATATTACAAAGG 1844
Db 401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg 420
QY 1845 GAGACAGAAGAAAGGGAAACAAAGAAAGAGAGAGAAAGAGAGAAAGAGGAGGAAAGA 1904
Db 421 GluThrGluGluGluLysGluGlnArgLysGluLysGluArgGluGluArg 440
QY 1905 CGTAATAACCGCAAGAGAGAATTTGACTAAGATCTTGGCTGAGTGGTTGAAGGGAA 1964
Db 441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys 460
QY 1965 AGCAATACGAAAGAGAGAGAGATTATAGGAAAATTAGGTCCAGGCCCTAGACAGTCAGGG 2024
Db 461 SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480
QY 2025 AACCTGGGCAATAGGACCCCACTCGACAGGACCAATTCGCATATTTGATAAAGAGAGGA 2084
Db 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500
QY 2085 CACTGGGCAAGAACTGCCCAAGAGGAAACAAAGAACCAAGGATTCCTAGCTCTAGAA 2144
Db 501 HisTrpAlaArgAsnCysProLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520
QY 2145 GAGCATAAAGATTAGGAGAGACGGGTTTCGACCCCTTCGACCCCGCCAGCGGTAACTTTG 2204
Db 521 GluAspLysAsp---GlyArgArgGlySerAspProLeuProGluProArgValThrLeu 539
QY 2205 AAGGTGAGGGGCGCAACCAAGTTGAGTTCTGTTGTTGATACCGGAGCGAAACATTCAGTGTA 2264
Db 540 LysValGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerValLeu 559

```

QY 2265 CTACAGCCATTAGGAAACTAAAGATATAAATAATCTCTGGTGATGGGTGCCACAGGSCAA 2324  
DB LeuGlnProLeuGlyLysLeuLysLeuLysLysSerTrpValMetGlyAlaThrGlyGln 579  
QY 2325 CAACAGTATCCATGGACTACCCGAGAACAGTTGACTTGGAGTGGAGCGGTAAACCCAC 2384  
DB ArgGlnTrpProTrpThrArgTrpValAspLeuGlyValGlyArgValThrHis 599  
QY 2385 TCGTTTCTGGTCTACACCTGAGTGCACACCCCTCTTAGTGAGACTTATTGACCAAG 2444  
DB SerPheLeuValIleProGluCysProValProLeuLeuGlyArgAspLeuLeuThrLys 619  
QY 2445 ATGGAGCACAATTTCTTTTGAACAAGGAAACAGAGTGTCTGCAATAACAACCT 2504  
DB MetGlyAlaGlnIleSerPheGlnGluGlnArgProGluValSerValAsnAsnLysPro 639  
QY 2505 ATCACTGTGTGACCTCCCAATTAGATGACGAATATCGACTATCTCTCCCTAGTAAG 2564  
DB IleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnValLys 659  
QY 2565 CCTGATCAAAATATACAATTTCTGTTGGAACAGTTTCCCAAGCCTGGGCAGAAACCGCA 2624  
DB ProAspGlnAspIleGlnSerTrpLeuGlnPheProGlnAlaTrpAlaGluThrAla 679  
QY 2625 GGGATGGTTTGGCAAGCAAGTTCCCCCAAGTTATTCAACTGAAGGCCAGTGCACCA 2684  
DB GlyMetGlyLeuAlaLysGlnValProProGlnValIleGlnLeuLysAlaSerAlaThr 699  
QY 2685 CCACTGCTCAGCAGACAGTACCTTGGTAAAGAGCTCAAGAGAAATTCGGCGCAT 2744  
DB ProValSerValArgGlnTyrProLeuSerArgGluAlaArgGluGlyIleTrpProHis 719  
QY 2745 GTCCAAAGATTAATCAACAGGCGATCCTAGTTCTCTGTCCTCAATCTCCCTGGAATCTCCC 2804  
DB ValGlnArgLeuIleGlnGlnIleLeuValProValGlnSerProTrpAsnThrPro 739  
QY 2805 CTGCTACCGGTTAGAAAGCTGGGACTTAATGACTATCGACCTACAGGACTTTGAGAGAG 2864  
DB LeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGlu 759  
QY 2865 GTCAATAACGGGTGAGGATATACACCCAGTCCCGAACCTTATAACCTTTGTGT 2924  
DB ValAsnLysArgValGlnAspIleHisProThrValProAsnProTyrAsnLeuLeuSer 779  
QY 2925 GCTCTCCCAACCCCAAGCTGATATACAGTATTGGACTTAAAGATGCTTCTCTGC 2984  
DB AlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeuLysAspAlaPheCys 799  
QY 2985 CTGAGATTACCCCCACTAGCCCAACCACTTTTTCCTTGAATGGAGATCCAGGTACG 3044  
DB LeuArgLeuHisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyThr 819  
QY 3045 GGAAGAACCGGCGAGCTCACCTGGACCCGACTCCCGAGGTTCAAGACTCCCGACC 3104  
DB GlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThr 839  
QY 3105 ATCTTTGACGAACCTTACACAGACCTGGCCAACTTCAGGATCAACACCTCTCAGGTG 3164  
DB IlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheLysIleGlnHisProGlnVal 859  
QY 3165 ACCCTCTCTCAGTACGTGGTACCTGCTTCTGGCGGAGCCCAACAGGACTGCTTA 3224  
DB ThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAlaThrLysGlnAspCysLeu 879  
QY 3225 GAAGGCAGCAGGCACTACTGCTGGAATTGCTCTGACTAGGCTACAGACCTCTGTAAG 3284  
DB GluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLys 899  
QY 3285 AGGGCCAGATTGAGGAGAGAGGTAAATACACTTGGGCTACAGTTTGGGACGGGAG 3344  
DB LysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGln 919

QY 3345 CGATGCTGACGAGGACGACGAAAGAAACTGTAGTCCAGATACCGGCCCCAACACACAGCC 3404  
DB ArgTrpLeuThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAla 939  
QY 3405 AAACAAATGAGAGACTTTTGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGTTT 3464  
DB LysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPhe 959  
QY 3465 GCACCTTAGCAGCCCACTCTACCCGCTAACCAAGAAAAAGGGAAATTTCTCTGGGCT 3524  
DB AlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGlyGluPheSerTrpAla 979  
QY 3525 CTGAGCAGCAGAGGATTTGATGTATCAAAAAGCCCTCTGTGAGCGCACCTGCTCTG 3584  
DB ProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAlaProAlaLeu 999  
QY 3585 GCCTCCCTGACCTAACTAACCCCTTACCTTTATGTGGATGAGCTGAGGAGTAGCC 3644  
DB AlaLeuProAspValThrLysProPheThrLeuTyrValAspGluArgLysGlyValAla 1019  
QY 3645 CGGGAGATTTTAAACCCAAACCTTAGGACCATGAGAGACCTCTCGCTACCTGTCAAAG 3704  
DB ArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLys 1039  
QY 3705 AAGCTCGATCCTGTACGAGTGGTTGGCCCATATGCTGAAGGCTATCGCAGCTGTGCC 3764  
DB LysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaValAla 1059  
QY 3765 ATACTGCTCAGGACGCTGACAAATTTGACTTTGGGACAGATATTAAGTAAATAGCCCC 3824  
DB IleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaPro 1079  
QY 3825 CATGCATTGAGAAACATCGTTCCGCGACCCCGAGCCGATGGATGACCAACGCCCGCATG 3884  
DB HisAlaLeuGluAsnIleValArgGlnProProAspArgTrpMetThrAsnAlaArgMet 1099  
QY 3885 ACCACTATCAAGCCTGCTTCTCACAGAGAGGTCACGTTCCCTCCACAGCCGCTCTC 3944  
DB ThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPheAlaProAlaAlaLeu 1119  
QY 3945 AACCTGCGCACTTCTGCTGAGAGACTGATGAACCACTGACTCATGATTCGCATCAA 4004  
DB AsnProAlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHisGln 1139  
QY 4005 CTATTGATTGAGGAGACTGGGTCGCCAAGGACTTTACAGACATACCGCTGACTGGAGAA 4064  
DB LeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThrGlyGlu 1159  
QY 4065 GTGCTAACCTGTTCACTCAACGGAAGCAGCTATGTGCTGGAAGGTAAAGGATGGCTGG 4124  
DB ValLeuThrTrpPheThrAspGlySerSerTyrValValGluGlyLysArgMetAlaGly 1179  
QY 4125 GGGCGGTGGTGGAGCGGACCGCAGATCTGGCCAGCAGCTGCGGAGGAACTTCA 4184  
DB AlaAlaValValAspGlyThrArgThrIleTrpAlaSerSerLeuProGluGlyThrSer 1199  
QY 4185 GCACAAAAGGCTGAGCTCATGCCCTCAGCAAGCTTTGCGGCTGGCCGAGGAAATCC 4244  
DB AlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSer 1219  
QY 4245 ATAAACATTATACGACAGCAGGTATGCTTTTGGACTGCGACACGCTACATGCGGCCATC 4304  
DB IleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIle 1239  
QY 4305 TATAACAAAGGGGTGTTGCTTACCTCAGCAGGAGGGAATTAAGAACAAAGAGGAATTT 4364  
DB TyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluIle 1259  
QY 4365 CTAAGCCTATTAGAGCCGCTACATTTACCAGAAAGCTAGCTATTATACACTGCTCTGGA 4424  
DB LeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIleIleHisCysProGly 1279  
QY 4425 CATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATGGCTGACCGGGTTGCCAAG 4484

Db 1280 HisGlnLysAlaLysAspLeuLysSerArgGlyAsnGlnMetAlaAspArgValAlaLys 1299  
Qy 4485 CAGCAGCCCGGGTGTAACTTCTGCTTATAATAGAAATGCCAAAGCCCGCAGAACCC 4544  
Db 1300 GlnAlaGlnAlaValAsnLeuLeuProIleGluThrProLysAlaProGluPro 1319  
Qy 4545 AGACGACGTACACCCCTAGAGACTGGCAAGATATAAAGATAGACCGATTCCTCTGAG 4604  
Db 1320 ArgArgGlnThrLeuGluAspTrpGlnGluLysIleAspGlnPheSerGlu 1339  
Qy 4605 ACTCCGAAGGAGCCTGCTATACCTCAGATGGGAAGAAATCTCCGCCCAACAAGAGGG 4664  
Db 1340 ThrProGluGlyThrCysTyThrSerTyGlyLysGluLeuLeuProHisLysGluGly 1359  
Qy 4665 TTAGAATATGTCACACAGATACATCGTCTAACCCACCTAGGAACATAACACCTGCAGCAG 4724  
Db 1360 LeuGluTyValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGln 1379  
Qy 4725 TTGGTCAAGACATCCCTTATCATGTCTGAGGCTACGAGAGTGGCTGACTCGGTGGTC 4784  
Db 1380 LeuValArgThrSerProTyHisValLeuArgLeuProGlyValAlaAspSerValVal 1399  
Qy 4785 AAAATATTGTGCTGCCAGCTGTTAAATGCTAATCTTCCAGAATGCCCTCCAGGGAAG 4844  
Db 1400 LysHisCysValProCysGlnLeuValAsnAlaAsnProSerArgIleProProGlyLys 1419  
Qy 4845 AGACTAAGGGAAACCCACCGCGCTCACTGGGAAGTGAATTCATCGAGGTAAAGCCG 4904  
Db 1420 ArgLeuArgGlySerHisProGlyAlaHisTrpGluValAspPheThrGluValLysPro 1439  
Qy 4905 GCTAAATACGAAACAATACCTATTGTTTCTGACACACCTTTCCAGGATGGTAGAG 4964  
Db 1440 AlaLysTyrglyAsnLysTyrglyLeuValPheValAspThrPheSerGlyTrpValGlu 1459  
Qy 4965 GCTTATCTCTAAGAAAGAGACTTCAACCGTGGTGGCTAAAAAATACTGGAAGAAATT 5024  
Db 1460 AlaTyrglyLysLysGluThrSerThrValValAlaLysLysIleLeuGluGluLe 1479  
Qy 5025 TTTCACAGATTGGAATACCTAAGTAAATAGGTGACAGCAATGTGTCAGCTTTTGTGCC 5084  
Db 1480 PheProArgPheGlyIleProLysValIleGlySerAspAsnGlyProAlaPheValAla 1499  
Qy 5085 CAGTGAAGTCAAGGACCTGGCCAGATATTGGGATTCATTTGGAACCTGATTCGTCATAC 5144  
Db 1500 GlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTy 1519  
Qy 5145 AGACCCCAAGCTCAGACACAGGTAGAGGATGAATAGAACCATTAAGAGACCTTTACT 5204  
Db 1520 ArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThr 1539  
Qy 5205 AAATTGACCGGAGACTGGCGTTAATGATGGATAGCTCTCTGCGCCTTTGTGCTTTT 5264  
Db 1540 LysLeuThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeuPhe 1559  
Qy 5265 AGGTTAGGACACCCCTGGACACTTTGGGCTGACCCCTATGATTAATTAATCTACCGGGA 5324  
Db 1560 ArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrglyLeuLeuLeuTyrgly 1579  
Qy 5325 CCCCCCATTTGGTAGAAATGCTTCTGTACATAGTGTGACGTGCTTCCAGCCT 5384  
Db 1580 ProProLeuValGluIleAlaSerValHisSerAlaAspValLeuLeuSerGlnPro 1599  
Qy 5385 TTGTTCTTAGGCTCAAGGCACTTGTAGTGGGTAGACAAACGAGCGTGAGGCAACTCCGG 5444  
Db 1600 LeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArg 1619  
Qy 5445 GAGGCTTACTCAGGAGGAGACTTCAGATCCCATCGTTTCCAAGTGGGAGATTCA 5504  
Db 1620 GluAlaTyrglyGlyGluGlyAspLeuGlnValProHisArgPheGlnValGlyAspSer 1639  
Qy 5505 GTCTACGTTAGACGCCACCGTGAGGAAACCTCCAGACTCGGTGGAGGGCCCTTATCTC 5564

Db 1640 ValTyrgValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProTyrgLeu 1659  
Qy 5565 GTACTTTTGACACACCAACGGCTGTGAAGTTCGAAGATCTCCACCTCGATCCATGCA 5624  
Db 1660 ValLeuLeuThrThrProThrAlaValLysValGluArgIleProThrTrpIleHisAla 1679  
Qy 5625 TCCACCTTTAAACCGCGCCACCTCCCGATTCGGGGTGGAAAGCCGAAAAAGACTGAAAAAT 5684  
Db 1680 SerHisValLysProAlaProProAspSerGlyTrpLysAlaGluLysThrGluAsn 1699  
Qy 5685 CCCTTAAAGCTTCGCTCCATCGGTGGTTCCTTACTCTCAATAAACCCTCT 5736  
Db 1700 ProLeuLysLeuArgLeuHisArgValValProTyrgSerValAsnAsn-SerSerSerMe 1719  
Qy 5736 ----- 5736  
Db 1719 tHisProThrLeuSerArgArgHisLeuProIleArgGlyLysProLysArgLeuLys 1739  
Qy 5737 -----CAGAC 5741  
Db 1739 sIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGlnVa 1759  
Qy 5742 TAATGTATCGCATAGGACAGACCTGAACTCCCATAAACCCCTTATCTCTCACCTGGTT 5801  
Db 1759 lAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrpLe 1779  
Qy 5802 AATTACTGACTCCGGCACAGGTATTATATCAACACACTCAAGGGAGGCTCCTTAGG 5861  
Db 1779 uLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeuGl 1799  
Qy 5862 AACCTGTGTGCTGCTACTATACCTTTCCTCAGATCAGTATTATCTCTAGTCTCTG 5913  
Db 1799 yThrTrpTrpProGluLeuTyrgValCysLeuArgSerValIleProGlyLeuAsnAspGl 1819  
Qy 5914 -ACCTCACCCCGATATCTCCATGCTCACGATTTATGTGTGCCCAGGACACCAAA 5972  
Db 1819 nAlaThrProProAspValLeuArgAlaTyrgLysPheTyrgValCysProGlyProProAs 1839  
Qy 5973 TAATGGAACATTTGGGAAATCCAGAGATTTCTTTGTAAACAATAGGACTGTGAAC 6032  
Db 1839 nAsnGluGluTyrgCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysIleTh 1859  
Qy 6033 CTCTAATGATGGATATTGGAAATGGCCAACTCTCAGCAGAGTAGGGTAAAGTTTCTTA 6092  
Db 1859 rSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrgPh 1879  
Qy 6093 TGTCACACCTATACCACTCTCGACAAATTAATTAC-----CTGACCTG 6137  
Db 1879 eValAsnAsnProThrSerTyrgAsnGlnPheAsnTyrgLysGlyArgTrpLysAspTr 1899  
Qy 6138 G-----ATTAGAACTGGNAGCCCAAGTCTCTCTTCAGACCT 6176  
Db 1899 pGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAspLe 1919  
Qy 6177 AGATTACTTAAATAAGTTTCTCTGAGAAAGGAAAAACAAGAAATATCTTAAATGGGT 6236  
Db 1919 uAspTyrgLysLysIleSerPheThrGluLysGlyLysGlnGlnIleGlnLysTrpVa 1939  
Qy 6237 AAATGTGTATCTTGGGAAATGGTATATTATGAGGCTCGGGTAAACAACACAGGCTCAT 6296  
Db 1939 lAsnGlyIleSerTrpGlyIleValTyrgLysGlySerGlyArgLysLysGlySerVa 1959  
Qy 6297 TCTAACTATTTCGCTCAAAATA---AACCACTGGAGCTCCATAGCTATAGACCAAA 6353  
Db 1959 lLeuThrIleArgLeuArgIleGluThrGlnMetGluProProValAlaIleGlyProAs 1979  
Qy 6354 TACGCTCTTACCGGTCAAAAGACCCCAACCCAA-----GGACAGGACCATC 6401  
Db 1979 nLysGlyLeuAlaGluGlnGlyProProIleGlnGlnArgProSerProAsnProSe 1999  
Qy 6402 CTCT---AACATACTTCTCGATCAGACCCCACTGAGCTTAACAGCAGCAGCTAAATGGG 6458  
Db 1999 rAspTyrgAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThrGl 2019

QY 6459 GGCAGAACTTTTACCTCATCCAGGAGAGCTTTTCAAGCTCTTAACTCCAGCTCCAGA 6518  
 DB 2019 YAlaLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAenSerThrThrProG 2039  
 QY 6519 GGCTACCTCTCTTGTGGTGTATGCTTGTGGGCCACCTTACTATGAAGAAATGCC 6578  
 DB 2039 uAlaThrSerSerCysTrpLeuAlaSerGlyProProTrpTrpGluGlyMetAl 2059  
 QY 6579 TAGAAGAGGAAATTCATGTGCAAAAGAACATAGAGACCAATGACATGGGATCCCA 6638  
 DB 2059 aArgGlyGlyPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerG 2079  
 QY 6639 AATAAGCTTACCTTACTGAGTTCCTGGAAAGGCACCTGCATAGAAAGTTCCTCC 6698  
 DB 2079 nAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyMetValPro 2099  
 QY 6699 ATCCCAACACACTTGTACCACTGAAAGCTTTAATCAAACTCTGAGAGTCAATA 6758  
 DB 2099 oSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerG 2119  
 QY 6759 TCTGGTACCTGGTTATCACAGGTGGTGGGCATGTAATCTGGATTAAACCCCTGTGTTTC 6818  
 DB 2119 rLeuValProGlyTyraAspArgTrpAlaCysAsnThrGlyLeuThrProCysValSe 2139  
 QY 6819 CACCTTGGTTTAAACCAACTAAAGATTTTGCATTATCGTCCAAATGTTCTCCCGAGT 6878  
 DB 2139 rThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArgVa 2159  
 QY 6879 GTATTACTACCCGAAAGCAATCTTGTGATGATATGATACATACAGAAATCATCGACAAA 6938  
 DB 2159 lTyTrpTrpProGluLysAlaValLeuAspLysTyraAspTrpArgTrpArgProLys 2179  
 QY 6939 GAGAGAACCATATCTCGACACTGTGTGATGCTCGAGTGGAGTGCAGCAGGTGT 6998  
 DB 2179 sArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVa 2199  
 QY 6999 AGGAACAGGAACAGTCCCTGTGTACGGACCAACAGCAGCTAGAACAGGACTTAGTAA 7058  
 DB 2199 lGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSerAs 2219  
 QY 7059 CCTACATCGAATGTACAGAACTCTCCAGCCCTAGAAATCTGTACGTAACCTGGA 7118  
 DB 2219 nLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuG 2239  
 QY 7119 GGAATCCCTAACCTCTTATCTGAAGTAGCTCTACAGAAATAGAGAGGGTTAGATTATT 7178  
 DB 2239 uGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLe 2259  
 QY 7179 ATTTCTAAAGAGGAGGATTTATGTAGCTTTGAAGAGGAAATGCTGTTTTTATGTGA 7238  
 DB 2259 uPheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysPheTyValAs 2279  
 QY 7239 TCATTACGGGCCATCAGACACTCCATGAACAAGCTTGAAGAAAGTGTGAGAAAGCTGC 7298  
 DB 2279 pHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArgArg 2299  
 QY 7299 AAGGAAAGAGAACTACTCAAGGTGGTTTGAAGGATGTTCAACAGGTCTCTTTGGTT 7358  
 DB 2299 sArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMe 2319  
 QY 7359 GGCTACCTTCTTCTTAAACAGGACCTTAAATAGTCTCTCTCTGTTACTCACAGT 7418  
 DB 2319 tThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrVa 2339  
 QY 7419 TGGGCCATGTATTATTAAACAGTAAATTTGCTTCTTATAGAACGAATAGTGCAGTCCA 7478  
 DB 2339 lGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaValG 2359  
 QY 7479 GATCATGGTACTTAGACAAACAGTACCAG 7506  
 DB 2359 nIleMetValLeuArgGlnGlnTrpGln 2369

## RESULT 2

Q9Q1X5\_9GAMR  
 ID Q9Q1X5\_9GAMR PRELIMINARY; PRT; 2376 AA.  
 AC Q9Q1X5;  
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DE Type C proviral gag, pol and env genes and LTR (class B, clone 33).  
 DE Porcine endogenous retrovirus.  
 OS Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
 OC Gammaretrovirus; 1-Mammalian type C virus group.  
 OX NCBI\_TaxID=61673;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Type C;  
 RX MEDLINE=20219394; PubMed=10756014;  
 RX DOI=10.1128/JVI.74.9.4028-4038.2000;  
 RA Czauderna F., Fischer N., Boller K., Kurth R., Toenjes R.R.;  
 RT "Establishment and characterization of molecular clones of porcine  
 RT endogenous retroviruses replicating on human cells.";  
 RL J. Virol. 74:4028-4038(2000).  
 DR EMBL; AJ133816; CAB65339.1; -; Genomic\_DNA.  
 DR HSP; P03385; IMOF.  
 DR SNR; Q9Q1X5; 7-98, 472-518, 2225-2277.  
 DR MEROPS; A02.020; -.  
 DR GO; GO:0019028; C:Viral capsid; IEA.  
 DR GO; GO:0004190; F:Aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0005198; P:DNA recombination; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO; GO:0019068; P:Viral assembly; IEA.  
 DR InterPro; IPR002050; Env\_Polyprotein.  
 DR InterPro; IPR008981; FmLVrecept-bind.  
 DR InterPro; IPR003036; Gag\_p30.  
 DR InterPro; IPR000840; G\_retro\_matrix.  
 DR InterPro; IPR001985; Peptidase\_A2\_cat.  
 DR InterPro; IPR001969; Pept\_Asp\_AS.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTse.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF01140; Gag\_MA; 1.  
 DR Pfam; PF02093; Gag\_p30; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR Pfam; PF00429; TLV\_coat; 1.  
 DR Pfam; PF00098; zf\_CCHC; 1.  
 DR SMART; SM00343; Znf\_C2HC; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS00879; RNASE\_H; 1.  
 DR PROSITE; PS0158; ZF\_CCHC; 1.  
 FT CHAIN 1 524 Gag protein.  
 FT CHAIN 525 1761 Pol protein.  
 FT CHAIN 1720 2376 Env protein.  
 SQ SEQUENCE 2376 AA; 266625 MW; 37DD39D3099DA4AF CRC64;

## Alignment Scores:

Pred. No.: 0 Length: 2376  
 Score: 11005.50 Matches: 2078  
 Percent Similarity: 91.5% Conservative: 99  
 Best Local Similarity: 87.3% Mismatches: 118  
 Query Match: 75.2% Indels: 84  
 DB: 2 Gaps: 12

US-10-723-552-3 (1-8132) x Q901X5_9GAMR (1-2376)		
QY	585	ATGGGACAGACGGGTGACGACCCCTCTTAGTTTGACTCTCGACCAATTGGACTGAAGTTAAA 644
Db	1	MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20
QY	645	TCCAGGCGTCATATTGTTCAGTTTCAGTTAAGAGGACCTTGGCAGACTTTCGTGTC 704
Db	21	SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysAla 40
QY	705	TCTGAATTGGCCGACATTTCGATGTGGATGGCCATTCAGAGGGGACCTTTAAATTCGAGATT 764
Db	41	SerGlnTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60
QY	765	ATCCTGGCTGTTAAGCAGATTATTTTCAGACTGGACCCGGCTCTCATCCGATCAGGAG 824
Db	61	IleLeuAlaValLysAlaIlePheGlnThrGlyProSerSerHisProAspGlnGlu 80
QY	825	CCCTATATCCTTACGTGGCAAGATTTCGCAGAGGATCCTCCGCCATCGGTTAAACCATGG 884
Db	81	ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100
QY	885	CTGAATAAGCCAAAGAACCCAGGTCCCGCAATTCTGGCTCTTGGAGAGAAAAACAAACAC 944
Db	101	LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120
QY	945	TCGGCTGAAAAGTCAAGCCCTCTCCTCATATCTACCCCGAGATTGAGGACCCACCGGCT 1004
Db	121	SerAlaGluLysValGluProSerProArgIleTyrProGluIleGluGluProProThr 140
QY	1005	TGGCCGGAAACCCCAATCTGTTCCCCACCCCTTATCTGGCACAGGTTGCCGAGGGGA 1064
Db	141	TrpProGluProGlnProValProProProTrpProAlaGlnGlyAlaValArgGly 160
QY	1065	CCCTTTGCCCTCTCTGGAGCTCCGCGGTGGAGGACCTGTCTGAGGAGACTCGGAGCCGG 1124
Db	161	ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180
QY	1125	AGGGGCGCACCCCGGAGCGGACAGACGAGATCCGACATTACCGCTGGCGACGTACGGC 1184
Db	181	ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly 200
QY	1185	CCTCCACACCGGGGGCCAAATTCAGCCCTCCAGTATTGGCCCTTTTCTCTGCAGAT 1244
Db	201	ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220
QY	1245	CTCTATAATTGGAAAACATAACATCCCTTTCTCGGAGGATCCCAACCGCTCACGGGG 1304
Db	221	LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240
QY	1305	TTGTTGAGTCCCTTATGTTCTCTCACGAGCTACTTGGGATGATTGTCACAGCTGCTG 1364
Db	241	LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
QY	1365	CAGACACTCTTCACAAACCGAGGACGAGAGAATTCTATTAGAGGCTAGAAAAAATGTT 1424
Db	261	GlnThrLeuPheThrThrGluLysGluArgIleLeuLeuGluAlaArgLysAsnVal 280
QY	1425	CCTGGGCGGACGGCGGACCCACCGGTGCAAAATGAGATTGATCGGATTTCCCTTA 1484
Db	281	ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300
QY	1485	ACTCGCCCGGTGGGACTACAAACCGCTGAAGGTAGGGAGACTTGAAAATCTATCGC 1544
Db	301	ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320
QY	1545	CAGGCTCTGTGGCGGGTCTCCGGCGGCGCTCAAGACGGCCCACTAAATTTGGCTAAGGTA 1604
Db	321	GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340
QY	1605	AGAGAAGTGATGACGAGGACCGAAATGAACCCCTCTGTGTTTTCTTGAGAGCTCTGGAA 1664
Db	341	ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
QY	1665	GCCTTCAGGCGGTACACCCCTTTTGATCCCACTCAGAGGCCCAAAAAGCCTCAGTGGCT 1724
Db	361	AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380
QY	1725	TTGGCCCTTTATAGACACAGTCAGCTCAGCTTGGATATTAGAAAAGACTTCAGAGACTGGAAGGG 1784
Db	381	LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400
QY	1785	TTACAGAGGCTGAGTTACCTGATCTAGTGAAGAGGACGAGAAAAGTATATTACAAAAGG 1844
Db	401	LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg 420
QY	1845	GAGCAGAACAAAGGGAACAAAGAAAGAGAGAGAAAGAGAGGAAAGGAGGAAGAAGA 1904
Db	421	GluThrGlnGluGlnLysGluGlnArgLysGluArgGluGluArgGluGluArg 440
QY	1905	CGTAATAAAACGGCAAGAGAGAATTTGACTTAAGATCTTGGCTGCAGTGTGAAGGAAA 1964
Db	441	ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValGluGlyLys 460
QY	1965	AGCAATACGGAAGAGAGAGAGATTTTAGGAAAATTAGTTCAGGCCCTAGACAGTCAGG 2024
Db	461	SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480
QY	2025	AACCTGGCAATAGGACCCCACTCGCAAGGACCAATGTGCATATTGTAAAGAAAGAGGA 2084
Db	481	AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500
QY	2085	CACCTGGCAAGGAACTCCGCCAAAGAGGAAACAAAGAGCAAGGATCTTAGCTCTAGAA 2144
Db	501	HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuThrLeuGlu 520
QY	2145	GAAGATAAAGATTAGGGGAGACGGGGTTTCGGACCCCTCCCGAGCCAGGGTTAACTTTG 2204
Db	521	GluAspLysAsp--GlyArgArgGlySerAspProLeuProGluProArgValThrLeu 539
QY	2205	AAGGTGGAGGGGCAACCGATTGATTCCTGTTGATACCGGCGGAACATTCAGTGCTA 2264
Db	540	LysValGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerValLeu 559
QY	2265	CTACAGCATTTAGGAAAACTAAAGATATAAAAAATCTCTGGGTGATGGGTGGCCACAGGCAA 2324
Db	560	LeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMetGlyAlaThrGlyGln 579
QY	2325	CACAGTATCCATGGACTACCCGAAAGAACAGTTGACTTTGGAGTGGGACGGTTAACCCAC 2384
Db	580	ArgGlnTyrProTrpThrThrArgArgThrValAspLeuAlaValGlyArgValThrHis 599
QY	2385	TCGTTTCTGTTGCTACCTGAGTGCACGACCCCTCTTAGTAGAGACTTATTACCAAG 2444
Db	600	SerPheLeuValIleProGluCysProValProLeuLeuGlyArgAspLeuLeuThrLys 619
QY	2445	ATGGGACACAAAATTTCTTTGAACAAAGGAAACCAAGAGTGTCTGCAAAATAACAAACCT 2504
Db	620	MetGlyAlaGlnIleSerPheGluGlnGlyArgProGluValSerValAsnAsnLysPro 639
QY	2505	ATCACTGTGTTGACCTCCAAATTAGATGACGAATATCGACTATATCTCTCCCTAGTAAAG 2564
Db	640	IleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnValLys 659
QY	2565	CCTGATCAAAATATACAATTTCTGTTGGAACAGTTTCCCAAGCTCGGCGAGAAACCGCA 2624
Db	660	ProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGlnAlaTrpAlaGluThrAla 679
QY	2625	GGGATGGGTTTGGCAAGCAAGTTCCCCCAAGTATTTCACACTGAAGGCCAGTCGCACA 2684
Db	680	GlyMetGlyLeuAlaLysGlnValProProGlnValIleGlnLeuLysAlaSerAlaThr 699
QY	2685	CCAGTGTCTAGACAGTACCTCCCTTGGTAAAGAGCTCAAGAGGAATTCGCCCGCAT 2744
Db	700	ProValSerValArgGlnTyrProLeuSerArgGluAlaArgGluGlyIleTrpProHis 719



QY 2745 GTCCAAAGATTAAATCCAAACAGGGCATCTAGTTCCTGCTCCAAATCTCCCTGGAATACTCCC 2804  
Db ValGlnArgLeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPro 739  
QY 2805 CTGCTACCGGTTAGAAAGCCCTGGCAGTAATGACTATGACCAAGTACAGGACTTCAGAGAG 2864  
Db LeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGlu 759  
QY 2865 GTCAATAAACCAGGTCAGGATATATACCCCAACAGCTCCCGAACCCCTTATAACCTCTTGTT 2924  
Db ValAsnLysArgValGlnAspIleHisProThrValProAsnProTyrAsnLeuLeuSer 779  
QY 2925 GCTCTCCACCCCAACGAGCTGTGTATACAGTATTGGACTTAAAGATGCTTCTTCGTC 2984  
Db AlaLeuProGluArgAsnTrpTyrThrValLeuAspLeuLysAspAlaPhePheCys 799  
QY 2985 CTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTTCGAATCGAGAGATCCAGGTACG 3044  
Db LeuArgLeuHisProThrSerGlnProLeuPheAlaPheGluTrpArgProGlyThr 819  
QY 3045 GGAAGAACCCGGGCAAGCTCACCTGACCCCGACTGCCCCCAAGGGTTCAAGAACTCCCCGACC 3104  
Db GlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThr 839  
QY 3105 ATCTTTGACGAGCCCTACACAGAGACTGGCCCAACTTCAGGATCCCAACACCTCAGGTG 3164  
Db IlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnVal 859  
QY 3165 ACCTCTCTCAGTACGTGGATGACCTGCTTCTGCGGGAGCCCAACAGGACTGCTTA 3224  
Db ThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAlaThrLysGlnAspCysLeu 879  
QY 3225 GAAGGCACGAGGCACTACTGCTGGAATTGCTGACCTAGCTACAGAGCTCTGCTTAAG 3284  
Db GluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLys 899  
QY 3285 AAGCCCAAGATTTCAGAGAGAGGTAACTACTTGGGGTACAGTTTGGCGGACGGCAG 3344  
Db LysAlaGlnIleCysArgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGln 919  
QY 3345 CGATGGCTGACGGGACCGGAGAAACTGTAGTCCAGATACCGGCCCAACACAGGCC 3404  
Db ArgTrpLeuThrGluAlaArgLysLysThrValValGlnIleProAlaProThrThrAla 939  
QY 3405 AAACAAATGACAGAGATTTTGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGTTT 3464  
Db LysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPhe 959  
QY 3465 GCGACCTTAGCAGCCCACTCTACCCGCTAACCAAGAAAAGGGGAATTTCTCTGGGCT 3524  
Db AlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGlyGluPheSerTrpAla 979  
QY 3525 CTTGAGCCACAGAGGATTTGATGCTATCAAAAGCCCTGCTGAGCGCACTGCTCTG 3584  
Db ProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAlaProAlaLeu 999  
QY 3585 GCCTCTCCTGACGTAACTAAACCTTTTACCCTTTATGTGGATGAGCGTAAGGAGTAGCC 3644  
Db AlaLeuProAspValThrLysProPheThrLeuTyrValAspGluArgLysGlyValAla 1019  
QY 3645 CGGGAGTTTAAACCCAAACCCCTAGGACCATGAGAGACCTGTGCGCTTACCTGTCAAAG 3704  
Db ArgGlyValLeuThrGlnThrLeuGlyProTrpArgProValAlaTyrLeuSerLys 1039  
QY 3705 AAGCTCGATCCTGTAGCAGTGTGTGGCCCATATGCTGAAGGCTATCGCAGCTGTGGCC 3764  
Db LysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaValAla 1059  
QY 3765 ATACTGCTCAAGCAGCTGACAAATTGACTTTTGGGACAGATATACTGTGTAATAGCCCC 3824  
Db IleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaPro 1079  
QY 3825 CATGCAATTGGAAACATCGTTCCGGCAGCCCCCAGACCGATGATGACCAACGCCCGCATG 3884

Db HisAlaLeuGluAsnIleValArgGlnProProAspArgTrpMetThrAsnAlaArgMet 1099  
QY 3885 ACCCACTATCAAAAGCCCTGCTTCTCACAGAGAGGTCAGTTCGCTCCACAGCCGCTCTC 3944  
Db ThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPheAlaProProAlaAlaLeu 1119  
QY 3945 AACCTCCCACTCTTCTGCTGAAAGACACTGATGAACAGTACTCATGATTGCCATCAA 4004  
Db AsnProAlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHisGln 1139  
QY 4005 CTATTGATTGAGAGACTGGGGTCCGCAAGACCTTACAGACATACACGCTGACTGGAGAA 4064  
Db LeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleProLeuThrGlyGlu 1159  
QY 4065 GTGCTAACCTGGTTCTACTGACGGAAGCAGCTATGTGCTGGAAGTAAGAGATCGCTGGG 4124  
Db ValLeuThrTrpPheThrAspGlySerTyrValValGluGlyLysArgMetAlaGly 1179  
QY 4125 GCGCGGTGTGGACGGGACCCGACGATCTGGGCCACAGACCTGCGCGAAGGAACCTTCA 4184  
Db AlaAlaValValAspGlyThrHisThrIleTrpAlaSerSerLeuProGluGlyThrSer 1199  
QY 4185 GCACAAAGGCTGAGCTCATGCGCCTCACGCAAGCTTTTGGCTGCGCCGAAGGAAATCC 4244  
Db AlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSer 1219  
QY 4245 ATAAACATTTATACGACACGACGATATGCTTTCGCTGCTGCACACGTACATCGGCGCATC 4304  
Db IleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIle 1239  
QY 4305 TATAACAAAGGGGTTGCTTACTCTCAGCAGGAGGAGAAATAAGAACAAAGAGAAAT 4364  
Db TyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluIle 1259  
QY 4365 CTAAGCCTATTAGAAGCCGTACATTTACCAAAAGGCTAGCTATTATACACTGTCCTGGA 4424  
Db LeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIleIleHisCysProGly 1279  
QY 4425 CATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATGCTGACCGGTTGCCAAG 4484  
Db HisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMetAlaAspArgValAlaLys 1299  
QY 4485 CAGCAGCCAGGGTGTAACTTCTGCTTATATAGAAATGCCCAAGCCCCAGAACCC 4544  
Db GlnAlaGlnAlaValAsnLeuLeuProIleIleGluThrProLysAlaProGluPro 1319  
QY 4545 AGACGACGTACACCTTAGAAGACTGGCAAGAGATAAAAAGATAGACCAAGTTCTCTGAG 4604  
Db ArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGlu 1339  
QY 4605 ACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGGAATCTGCCCAACCAAGAGGG 4664  
Db ThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGly 1359  
QY 4665 TTAGAATATGTCACACAGATACATCGTCTAACCCACTAGAACTAAACACCTGCACAG 4724  
Db LeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGln 1379  
QY 4725 TTGGTCAGAACATCCCTTATCATGTTCTGAGGCTACCAAGAGTGGCTGACTCGGTGTC 4784  
Db LeuValArgThrSerProTyrHisValLeuArgLeuProGlyValAlaAspSerValVal 1399  
QY 4785 AAACATTGTGTCCTCCAGCTGGTGTAACTGCTTAATCCTTCCAGATGCTCCAGGGAAG 4844  
Db LysHisCysValProCysGlnLeuValAsnAlaAsnProSerArgIleProProGlyLys 1419  
QY 4845 AGACTAAGGGGAGGCCACCCAGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCG 4904  
Db ArgLeuArgGlySerHisProGlyAlaHisTrpGluValAspPheThrGluValLysPro 1439  
QY 4905 GCTAAATACGGAAACAAATACCTATTGGTTTTGTAGACACCTTTTTCAGGATCGGTAG 4964

Db 1440 AlalysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpValGlu 1459  
Qy 4965 GCTTATCTACTAAGAAGACATTCACCGTGTGGCTAAATAAATACTCGGAAGAAATT 5024  
Db 1460 AlaTyrProThrLysGluThrSerThrValValAlaLysLysIleLeuGluGluIle 1479  
Qy 5025 TTTCCAAGATTGGAATACCTTAAGCTAATAGGTCAGACAATGGTCCAGCTTTTGTGGCC 5084  
Db 1480 PheProArgPheGlyIleProLysValIleGlySerAspAsnGlyProAlaPheValAla 1499  
Qy 5085 CAGTAAGTCAGGACCTGGCCAGCATATTGGGATGATTGGAACTGCATTGTCATAC 5144  
Db 1500 GlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyr 1519  
Qy 5145 AGACCCAAAGCTCAGACAGGTAGAGAGATGAATAGAACCATTAAGAGACCTTACT 5204  
Db 1520 ArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThr 1539  
Qy 5205 AAATTGACCGGAGACTGGCGTTAATGATTGGATAGCTCTCCGCTTGTGCTTTT 5264  
Db 1540 LysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeuLeuProPheValLeuPhe 1559  
Qy 5265 AGGTTAGGAACACCCCTGGACAGCTTTGGGTGACCCCTATGAATTACTCTACGGGGGA 5324  
Db 1560 ArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGluLeuLeuTyrGly 1579  
Qy 5325 CCCCCCGATTGGTAAATGCTTCTGTACATAGTCTGACGTGCTTCCAGCCT 5384  
Db 1580 ProProLeuValGluIleAlaSerValHisSerAlaAspValLeuLeuSerGlnPro 1599  
Qy 5385 TTGTTCTCTAGGCTCAAGGCATTGAGTGGGTGAGACAACGAGCGTGGAGCACTCCGG 5444  
Db 1600 LeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArg 1619  
Qy 5445 GAGCCTACTCAGGAGGAGACTTCGAGATCCACATCGTTTCCAGTGGGAGATTCA 5504  
Db 1620 GluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArgPheGlnValGlyAspSer 1639  
Qy 5505 GTCTACGTTAGACGCCACCGTGAGGAACCTCAGACTCGGTGGAAGGCCCTTATCTC 5564  
Db 1640 ValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProTyrLeu 1659  
Qy 5565 GTACTTTTGACCAACACCGCTGTCAAGTCCAAAGAAATCTCCACTGGATCCATGCA 5624  
Db 1660 ValLeuLeuThrThrProThrAlaValLysValGluGlyIleSerThrTrpIleHisThr 1679  
Qy 5625 TCCACAGTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAGCCGAAAGACTGAAAT 5684  
Db 1680 SerHisValLysLeuAlaProProAspSerGlyTrpArgAlaGluLysThrGluAsn 1699  
Qy 5685 CCCCTTAAGCTTCGCCCTCCATCGCGTGGTTCCTTACTCTGTCAATAACCTCT 5736  
Db 1700 ProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsnAsnAsn-SerProGlyG 1719  
Qy 5736 ----- 5736  
Db 1719 nIleHisProThrLeuSerTrpArgHisLeuLeuProThrArgGlyGluProLysArgLe 1739  
Qy 5737 -----CA 5738  
Db 1739 uArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrIleThrProG 1759  
Qy 5739 GACTAATGGTATGGCATAGAGAGACGCTGAATCCCATAAACCTTATCTCTCAGCTG 5798  
Db 1759 nAlaSerSerLysArgLeuIleAspSerSerAsnProHisArgProLeuSerLeuThrTr 1779  
Qy 5799 GTTAATTACTGACCCGCGACAGGTATTAAATATCAACAACACTCAAGGGGAGGCTCTTT 5858  
Db 1779 pLeuIleLeuAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProAr 1799  
Qy 5859 AGGAACCTGGTGGCTATCTATACGTTTGGCTCAGATCAGTTATTCAGTCTGACCTC 5918  
Db 1799 gGlyThrTrpTrpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSe 1819

Qy 5919 A----CCCCAGATATCTCCTCATCTCAGGATTTTATGTTTGGCCAGGACCCACCAATAA 5975  
Db 1819 rThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysG 1839  
Qy 5976 TGGAAACATTGGGGAATCCACAGAGATTTCTTTGTGTAACAATGGAACTGTGTAACTC 6035  
Db 1839 u---LysTyr-CysGlyAspSerGlyGluSerPheCysArgArgTrpSerCysValThrSe 1858  
Qy 6036 TAATGATGGATATTGGAAATGGCAACCTCTCAGCAGGATAGGTAAGTTTCTTATGT 6095  
Db 1858 rAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVa 1878  
Qy 6096 CAACACCTATACCAGCTCTCGACAAATTTAATTACTGACCTGGATTAGAGCTGAAGCCC 6155  
Db 1878 lAsn-----SerGlyProGlyLysTyrLysValMetLysLeuTyrLysAspLy 1894  
Qy 6156 CAAGTGTCTCTCCTTACAGACCTAGATTACCTAAAAATAAGTTTTCACCTGAGAAAGAAACA 6215  
Db 1894 sSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysG 1914  
Qy 6216 AGAAATATCTTAATGGGTAATGGTATGCTCTGGGGAATGGTATAT-----TATGG 6269  
Db 1914 nGluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTyrLysTyrG 1934  
Qy 6270 AGGCTCGGTTAAACAACCGCTCCATTCTTAACCTATTTCGCTCAAAATAAACCCAG---CT 6326  
Db 1934 yGlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyTh 1951  
Qy 6327 GAGGCTTCAATGGCTATAGGACCAATACGCTCTTACGCGGTCAAAAGACCCCA----- 6381  
Db 1951 rGluProProValAlaValGlyProAspLysValLeuAlaGluGlnGlyProProAlaLe 1971  
Qy 6382 -----ACCA 6386  
Db 1971 uGluProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrG 1991  
Qy 6387 AGGACACGAGGACCTCTCTTAACATACTCTCGATCAGACCCCACTCAGTCT-----AA 6440  
Db 1991 nSerPro-----SerAsnGlyThrThrGlyLeuIleProThrAsnThrProArgAs 2008  
Qy 6441 CAGC-----ACGACTAAATGGGGCAAAATCTTTTAGCCTCATCCAGGAGCTTT 6491  
Db 2008 nSerProGlyValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPh 2028  
Qy 6492 TCAAGCTCTTAACCTCAGGACTCCAGAGGCTACTCTTCTTGTGGTATGCTTAGCTTC 6551  
Db 2028 eGlnAlaIleAsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSe 2048  
Qy 6552 GGGCCACCTTACTATCAAGGAATGGCTAGAAGGGGAAATTTCAATGTGACAAAGAACA 6611  
Db 2048 rGlyProProTyrTrpGluGlyMetAlaLysGluGlyLysPheAsnValThrLysGluHi 2068  
Qy 6612 TAGAGACCAATGACATGGGGATCCCAAAATAAGCTTTACCCCTTACTGAGGTTTCTGAAA 6671  
Db 2068 sArgAsnGlnCysThrTrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLy 2088  
Qy 6672 AGGACCTGTCATAGGAAGGTTTCCCATCCCAACACACTTTGTGTAAACACACTGAAGC 6731  
Db 2088 sGlyThrCysIleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValVa 2108  
Qy 6732 CTTTAATCAACCTCTCAGAGTCAATATCTGGTACCTGGTTATGCAGAGTGGTGGCATG 6791  
Db 2108 lTyrGluGlnAlaSerGlnAsnGlnIleValProGlyTyrAsnArgTrpIleAlaCy 2128  
Qy 6792 TAATACTGGATTAAACCCCTTGTGTTCACCTTGGTTTTTAACCAAACTAAAGATTTTTG 6851  
Db 2128 sAsnThrGlyLeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheCy 2148  
Qy 6852 CATTATGGTCAAAATTTGTTCCCGAGTGATTACTATCCGAAAAAGCAATCCTTGTATGA 6911  
Db 2148 sValMetValGlnIleValProArgValTyrTyrHisProGluGluValValLeuAspG 2168

```
QY 6912 ATATGACTACAGAAATCATCCACAAAAGAGAGACCCCATATCTCTGACACTTGTCTGTGAT 6971
Db 2168 UTYrAspTYrArgTYrAsnArgProLeuArgGluProValSerLeuThrLeuAlaValLe 2188
QY 6972 GCTCGGACTCGAGTGGCAGCAGGTGTAGGAACAGGACAGCTGCCCTGGTCACGGGACC 7031
Db 2188 tLeuGlyLeuGlyThrAlaValGlyValGlyThrAlaAlaLeuIleThrGlyPr 2208
QY 7032 ACAGCAGCTAGAAAACAGCAGCTTAGTAACTCATCGAATTCTTAACAGAGAATCTCCAAGC 7091
Db 2208 oGlnGlnLeuGluGlyLeuGlyLeuGlyLeuHisAlaAlaMetThrGluAspLeuArgAl 2228
QY 7092 CCTAGAAAATCTGTCACTACCTCGGAGGATCCCTAACCTCTTATCTTGAAGTAGTCCT 7151
Db 2228 aLeuGluGluSerValSerAsnLeuGluSerLeuThrSerLeuSerGluValValLe 2248
QY 7152 ACAGATAGAGAGAGGTTAGATTATTATTCTTAAAGAGAGGAGGATTTATGTAGCCTT 7211
Db 2248 uGlnAsnArgGlyLeuAspLeuLeuPheLeuArgGlyGlyLeuCysAlaAlaLe 2268
QY 7212 GAAGGAGGATGCTGTTTTTATGTGGATCATTCAGGGGCCATCAGAGACTCCATGAACAA 7271
Db 2268 uLysGluGluCysCysPheTYrValAspHisSerGlyAlaIleArgAspSerMetSerLy 2288
QY 7272 GCTTAGAGAAAGTTGGAGAGCGTCGAGGAGAAAGAAACTACTCAAGGCTGTTGA 7331
Db 2288 sLeuArgGluArgLeuGluArgArgArgGluArgGluAlaAspGlnGlyTrpPheGl 2308
QY 7332 GGGATGTTCAACAGGCTCTCTTGGTGGCTACCTACTTCTGCTTTACAGGACCTT 7391
Db 2308 uGlyTrpPheAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLe 2328
QY 7392 AATAGTCTCTCTCTGTTACTCACAGTGGGCCATGTATTATTAAACAAGTTAATGCTT 7451
Db 2328 uValValLeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPh 2348
QY 7452 CATTAGAGAACGAATAGTCAGTCCAGATCATGCTACTGATTAGACAAAGATACCAA 7506
Db 2348 eValArgGluArgValSerAlaValGlnIleMetValLeuArgGlnGlnTrpGln 2366

RESULT 3
OQ91X3_9GAMR
ID OQ91X3_9GAMR PRELIMINARY; PRT; 2376 AA.
AC OQ91X3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Type C proviral gag, pol and env genes and LTR (class B, clone
DE 43).
OS Porcine endogenous retrovirus.
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC Gammaretrovirus; 1-Mammalian type C virus group.
OX NCBI_TaxID=61673;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Type C;
RX MEDLINE=20219394; PubMed=10756014;
RX DOI=10.1128/JVI.74.9.4028-4038.2000;
RA Crauderna F., Fischer N., Boller K., Kurth R., Toenjes R.R.;
RT "Establishment and characterization of molecular clones of porcine
RL endogenous retroviruses replicating on human cells.";
RL J. Virol. 74:4028-4038(2000).
DR EMBL; AJ133818; CAB65341.1; -; Genomic_DNA.
DR HSSP; P03385; IMOF.
DR SMR; OQ91X3; 7-98, 2225-2277.
DR MEROPS; A02.020; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
```

```
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0019068; P:viral assembly; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; FmuLvrecept-bind.
DR InterPro; IPR003036; Gag_p30.
DR InterPro; IPR000840; G_retro.matrix.
DR InterPro; IPR001995; Peptidase_A2_cat.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVse.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01140; Gag_MA; 1.
DR Pfam; PF02093; Gag_p30; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT_1; 1.
DR Pfam; PF00429; TLV_coat; 1.
DR Pfam; PF00098; zf_CCHC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS01175; ASP_PROT_RETROV; 1.
DR PROSITE; PS00879; RNASE_H; 1.
DR PROSITE; PS01158; ZF_CCHC; 1.
FT CHAIN 1 524 Gag protein.
FT CHAIN 525 1761 Pol protein.
FT CHAIN 1720 2376 Env protein.
SQ SEQUENCE 2376 AA; 266550 MW; C6D16E368BA49AEF CRC64;

Alignment Scores:
Pred. No.: 0 Length: 2376
Score: 10997.50 Matches: 2074
Percent Similarity: 91.5% Conservative: 101
Best Local Similarity: 87.3% Mismatches: 123
Query Match: 75.1% Indels: 78
DB: 2 Gaps: 11

US-10-723-552-3 (1-8132) x OQ91X3_9GAMR (1-2376)
QY 585 ATGGGACAGACGGTGACGACCCCTCTTAGTTTGACTCTCGACCATTCGACTGAAGTTAAA 644
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20
QY 645 TCCAGGGCTCATATTTGTCAGTTCAGTTAAGAGGACCTTCGGCAGACTTCTGTGTC 704
Db 21 SerArgAlaHisAsnLeuSerValGlnValLysGlyProTrpGlnThrPheCysAla 40
QY 705 TCTGAATGGCCGACATTCGATGTTGGATGCCCATCAGAGGGGACCTTTAATTCTGAGATT 764
Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60
QY 765 ATCCTGGCTGTTAAAGCAGTTATTTTTCAGACTGGACCCCGCTCTCATCCGATCAGGAG 824
Db 61 IleLeuAlaValLysAlaIleIlePheGlnThrGlyProSerSerHisProAspGlnGlu 80
QY 825 CCCTATATCTTACGTGGCGAAGATTTGGCAGAGGATCCTCCGCCATGGGTTAAACCATGG 884
Db 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100
QY 885 CTGAATAAGCAAGAAGCCAGCTCCCGCAATTTCTGCTCTCGAGAGAGAAAACAAACAC 944
Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGlyAsnLysHis 120
QY 945 TCGGCTGAAAAGCTCAAGCCCTCTCTCTCATATCTACCCCGAGATTGAGGAGCCACCGGCT 1004
Db 121 SerAlaGluLysValGluProSerProArgIleTyProGluIleGluGluProProThr 140
QY 1005 TGCCCGGAACCCCAATCTGTTTCCCAACCCCTTATCTGCGACAGGGTCCGGGAGGGGA 1064
Db 141 TrpProGluProGlnProValProProProTyProAlaGlnGlyAlaValArgGly 160
```

1065 CCCTTTGCCCTCTGGAGCTCCGGCGTGAGGAGCCTGCTGCAGGACTCGGAGCGG 1124  
Db |||||  
161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180  
Qy AGGGCGCCACCCCGGAGCGACAGACAGATCGCATTCACCGCTCGCGCACCTAGCGC 1184  
Db |||||  
181 ArgGlyAlaThrProGluArgThrAspGluLeuAlaLeuProLeuArgThrTyrGly 200  
Qy CCTCCACACCGGGGGGCAATGTCAGCCCTCCAGATATTCGCTTTCCTTCGCAGAT 1244  
Db |||||  
201 ProProMetProGlyGlyGlnLeuProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
Qy CTCTATAATTCGAAAACTAACCATCCCTCTTCGGAGGATCCCAACGCTCACCGGG 1304  
Db |||||  
221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
Qy TTGCTGGAGTCCCTTATCTCTCACCGAGCTACTTGGGATGATTGTCAACAGCTGCTG 1364  
Db |||||  
241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnLeuLeu 260  
Qy CAGACACTCTTCACACCGAGAGCGAGAGAAATTCATTAGAGGCTAGAAAAATGTT 1424  
Db |||||  
261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280  
Qy CCTGGGCGGACGGCGGACCCACCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA 1484  
Db |||||  
281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
Qy ACTGCGCCGGTGGGACTACAAACACGCTGAAGTAGGAGAGCTTGAAAATCTATCGC 1544  
Db |||||  
301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
Qy CAGGCTCTGTGGCGGGTCTCGGGGCCCTCAAGCGGCCACTAATTTGGCTAAAGTA 1604  
Db |||||  
321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
Qy AGAGAGTGATGCAGGGACCAANTGMACCCCCTCTGTTTTCTTGAGAGCTCTTGAA 1664  
Db |||||  
341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
Qy GCCTTCAGCGGTACACCCCTTTTGATCCCACTCAGAGGCCCAAAAGCCTCAGTGGCT 1724  
Db |||||  
361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
Qy TTGGCTTTTATAGGACAGCTCAGCCTTGATATTTAGAAAGAGCTTCAGAGACTGGAAGG 1784  
Db |||||  
381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
Qy TTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGCGAGAGAAAGTATATTACAAAAG 1844  
Db |||||  
401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg 420  
Qy GAGACAGAAGAAGAAAGGAAACAAAGAAAGAGAGAGAAAGAGAAAGGAGGAAAGA 1904  
Db |||||  
421 GluThrGluGluGluLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 440  
Qy CGTAATAAACCGCGAAGAGAAGAAATTTGACTAAGATCTTGGCTGCGAGCTGGTTGAAGGAAA 1964  
Db |||||  
441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValValGluGlyLys 460  
Qy AGCAATACGGAAGAGAGAGATTTTAGGAAATTAGTGCAGGCCCTCAGACAGTCAGG 2024  
Db |||||  
461 SerSerArgGluArgGluAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
Qy AACCTGGGCAATAGAGCCCACTGCACAAGACCAATGTGCATATTTGTAAGAAGAGAGA 2084  
Db |||||  
481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500  
Qy CACTGGGCAAGAACTGCCCCAAGAGGAAACAAAGCAACCAAGATCTTAGCTCTAGAA 2144  
Db |||||  
501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520

2145 GAAGATAAAGATTAGGGAGACGGGTTTCGAGCCCTCCCGAGCCAGGGTAACCTTGT 2204  
Db |||||  
521 GluAspLysAsp---GlyArgArgGlySerAspProLeuProGluProArgValAlaLeu 539  
Qy AAGTGGAGGGCAACACAGTTGAGTTCTCGTGTGATACCGGAGCGAAACATTCTAGTGCTA 2264  
Db |||||  
540 LysValGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerValLeu 559  
Qy CTACAGCCATTAGGAAAACATAAAGATAAAAAATCTGGGTGATGGGTGCCACAGGGCAA 2324  
Db |||||  
560 LeuGlnProLeuGlyLysLeuLysLysSerTrpValMetGlyAlaThrGlyGln 579  
Qy CAACAGTATCCATCGACTACCCGAAGAACACTTTCACCTTGGAGTGGGACGGGTAAACCCAC 2384  
Db |||||  
580 ArgGlnTyrProTrpThrThrArgArgThrValAspLeuAlaValGlyArgValThrHis 599  
Qy TCGTTTCTGGTCATPACTGAGTGGCCAGCACCCCTCTTTCAGTAGAGACTTATTGACCAAG 2444  
Db |||||  
600 SerPheLeuValIleProGluCysProValProLeuLeuGlyArgAspLeuLeuThrLys 619  
Qy ATCGGAGCACAAATTTCTTTTGACAAAGGGAACACAGAAGTCTGCAAAATAACAAACCT 2504  
Db |||||  
620 MetGlyAlaGlnIleSerPheGluGlnGlyArgProGluValSerValAsnAsnLysPro 639  
Qy ATCACTGTGTGACCCCTCCCAATTAGATGACCAATATCGACTATACTCTCCCTTAGTAAAG 2564  
Db |||||  
640 IleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnValLys 659  
Qy CCTGATCAAAATATPACAATTCTGTTGGAACAGATTTCCCAAGCTGGGCGAGAAACCGCA 2624  
Db |||||  
660 ProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGlnAlaTrpAlaGluThrAla 679  
Qy GGGATGGGTTGGCAAAAGCAAGTTCCCCCAAGTATTCAACTGAGGCCAGTGCACACA 2684  
Db |||||  
680 GlyMetGlyLeuAlaLysGlnValProProGlnValIleGlnLeuLysAlaSerAlaThr 699  
Qy CCAGTGTCACTCAGACAGTACCCCTTCAGTAAAGAAAGCTCAAGAGGAATTCGGCCGAT 2744  
Db |||||  
700 ProValSerValArgGlnTyrProLeuSerArgGluAlaArgGluGlyIleTrpProHis 719  
Qy GTCCAAAGATTAAATCCCAACAGGGCATCTAGTTCTCTGTCCTTTCCTGGAATCTCCCTGAAATCTCCC 2804  
Db |||||  
720 ValGlnArgLeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPro 739  
Qy CTGTACTCGGTAGAAAGCCCTGGCACTAATGACTATCGACCAGTACAGGACTTCGAGAG 2864  
Db |||||  
740 LeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGlu 759  
Qy GTCAATAAACCGGTGCGAGGATATACACCAACAGTCCCGAACCCCTTATAACCTCTTGTGT 2924  
Db |||||  
760 ValAsnLysArgValGlnAspIleHisProThrValProAsnProTyrAsnLeuLeuSer 779  
Qy GCTCTCCACCCCAACCGGAGCTGTATACAGTATTGACTTAAAGGATGCTCTTCTGCG 2984  
Db |||||  
780 AlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeuLysAspAlaPhePheCys 799  
Qy CTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTTCGAATCGAGAGATCCAGGTACG 3044  
Db |||||  
800 LeuArgLeuHisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyThr 819  
Qy GGAAGAACCGGGCAGCTCACCTGGACCCGAGCTGCCCAAGGGTTCAAGAACTCCCGCACC 3104  
Db |||||  
820 GlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThr 839  
Qy ATCTTTGACGAAGCCCTACACAGAGACTGGCCCACTTCAGGATCCCAACACCTCAGGTG 3164  
Db |||||  
840 IlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnVal 859  
Qy ACCCTCTCCAGTACGTGGATGACTCTGCTGGGGGAGCCACCAACAGACTGCTTA 3224  
Db |||||  
860 ThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAlaThrLysGlnAspCysLeu 879  
Qy GAAGGCACGAAGGCACCTACTCTCGAATTGTCTGACCTTAGGCTACAGAGCCTCTGCTAAG 3284

Db 880 GluGlyThrLysAlaLeuLeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLys 899  
QY 3285 AAGGCCCAAGATTTCAGGAGAGAGTAACTATGTTGGGTACAGTTTGGCGGACGGGCGAG 3344  
Db 900 LysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyrSerLeuArgGlyGln 919  
QY 3345 CGATGGCTGACGGAGGACCGAAGAAACTGTAGTCCAGATACCGGCCCCCAACACAGCC 3404  
Db 920 ArgTrpLeuThrGluAlaArgLysThrValValGlnIleProAlaProThrThrAla 939  
QY 3405 AAACAAATGAGAGATTTCGGGACACAGCTGGATTTCAGACTGTGGATCCCGGGTTT 3464  
Db 940 LysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPhe 959  
QY 3465 GCGACCTTAGCAGCCCACTTACCCGCTAACCAAGAAAAGGGGAATTCTCTCGGCT 3524  
Db 960 AlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGlyGluPheSerTrpAla 979  
QY 3525 CCTGAGCACCAAGGCAATTGTATGCTATCAAAAGCCCTGTGAGCGCACCTGTCTGT 3584  
Db 980 ProGluHisGlnLysThrPheAspAlaIleLysLysAlaLeuLeuSerAlaProAlaLeu 999  
QY 3585 GCCCTCCTGACGTAACTAAACCTTTACCTTTATGTGGATGAGCTAAGGAGTAGCC 3644  
Db 1000 AlaLeuProAspValThrLysProPheThrLeuTyrValAspGluArgLysGlyValAla 1019  
QY 3645 CGGGAGCTTTTAAACCAACCTAGGACCATGGAGAAGACCTGTCGCTACCTGTCAAAG 3704  
Db 1020 ArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLys 1039  
QY 3705 AAGCTCGATCCTGTAGCAGTGGTGGCCCATATGCTGAAGGCTATCGCAGCTGTGGCC 3764  
Db 1040 LysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaValAla 1059  
QY 3765 ATACTGTCAAGACGCTGACAAATTGACTTTGGGACAGATATACTGTAATAGCCCCC 3824  
Db 1060 IleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaPro 1079  
QY 3825 CATGCTTGGAGAACATCGTTTCGCGACGCCCGGACGATGGATGACCAACGCCCGCATG 3884  
Db 1080 HisAlaLeuGluAsnIleValAlaGlnProAspArgTrpMetThrAsnAlaArgMet 1099  
QY 3885 ACCCACTATCAAGCCCTCTTCACAGAGGGTCACTGCTCGCTCCACAGCGCTCTC 3944  
Db 1100 ThrHisTyrGlnSerLeuLeuThrGluArgValThrPheAlaProProAlaAlaLeu 1119  
QY 3945 AACCTGCCACTCTTCGCTGAAGACACTGATGAACCACTGACTCATGATTCGCATCAA 4004  
Db 1120 AsnProAlaThrLeuLeuProGluGluThrAspGluProValThrHisAspCysHisGln 1139  
QY 4005 CTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAGAA 4064  
Db 1140 LeuLeuIleGluThrGlyValArgLysAspLeuThrAspIleProLeuThrGlyGlu 1159  
QY 4065 GTGCTAACCTGGTTTCACTGACGGAAGCAGCTATGTGTGGAAGGTAAAGGATGGTGGG 4124  
Db 1160 ValLeuThrTrpPheThrAspGlySerSerTyrValValGluGlyLysArgMetAlaGly 1179  
QY 4125 GCGCGGTGTGACGGGACCGGACGATCTGGGCCAGCAGCTGCGCGGAAGCACTTCA 4184  
Db 1180 AlaAlaValValAspGlyThrHisThrIleTrpAlaSerSerLeuProGluGlyThrSer 1199  
QY 4185 GCACAAAGGCTGAGCTCATGGCCCTCACCGAGCTTTGGCTGGCCGAGGGAATCC 4244  
Db 1200 AlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSer 1219  
QY 4245 ATAAACATTATACGACACAGGTATGCTTTGGACTCCACACGCTACATGATGGGCCATC 4304  
Db 1220 IleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIle 1239  
QY 4305 TATAAACAAAGGGGTGTCTTACTCAGCAGGAGGGAATAAAGAACAAAGAGGAAATT 4364

Db 1240 TyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluIle 1259  
QY 4365 CTAAGCCTATTAGAAGCGGTACATTTTACCAAAAGGCTAGCTATTATATACACTGTCCTGGA 4424  
Db 1260 LeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIleIleHisCysProGly 1279  
QY 4425 CATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATGGCTGACCGGTTGCCAAG 4484  
Db 1280 HisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMetAlaAspArgValAlaLys 1299  
QY 4485 CAGGACCCCGGGTGTAAACCTTCTGCCTATTAATAGAAATGCCCAAGCCCGCAAGACCC 4544  
Db 1300 GlnAlaAlaGlnAlaValAsnLeuLeuProIleIleGluThrProLysAlaProGluPro 1319  
QY 4545 AGACGACAGTACACCCCTAGAAAGACTGGCAAGAGATAAAAGATAGACACCTCTCTGAG 4604  
Db 1320 ArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGlu 1339  
QY 4605 ACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATCTCTGCCCAACAAGAGGG 4664  
Db 1340 ThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGly 1359  
QY 4665 TTAGATATATGTCACACGATACATCGTCTAACCCCTAGAACTAAACACCTGCGACAG 4724  
Db 1360 LeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGln 1379  
QY 4725 TTGCTCAGAACATCCCTTATCATGTTCTCAGAGCTACCGAGTGGCTGACTCGGTGTC 4784  
Db 1380 LeuValArgThrSerProTyrHisValLeuArgLeuProGlyValAlaAspSerValVal 1399  
QY 4785 AAACATTGTGTGCCCTGCCAGCTGGTTAATGCTTAATCTTCCAGAAATGCTCCAGGGAAG 4844  
Db 1400 LysHisCysValProCysGlnLeuValAsnAlaAsnProSerArgValProProGlyLys 1419  
QY 4845 AGACTAAGGGGAAGCACCCAGCGCTCACTGGGAAGTGGACTTCACTAGGTAAAGCCG 4904  
Db 1420 ArgLeuArgGlySerHisProGlyAlaHisTrpGluValAspPheThrGluValLysPro 1439  
QY 4905 GCTAAATACGGAACAAATACCTATTGTTTGTAGACACCTTTTTCAGATGGGTAGAG 4964  
Db 1440 AlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpValGlu 1459  
QY 4965 GCTTATCTACTAAGAAAGAGACTTCAACCGTGGTGGCTTAAATACTTAAAGAAATTT 5024  
Db 1460 AlaTyrProThrLysLysGluThrSerThrValValAlaLysLysIleLeuGluIle 1479  
QY 5025 TTTCCAAGATTGGAATACCTAAGGTAAATAGGTGACAAATGGTCAGCACTTTGTTGCC 5084  
Db 1480 PheProArgPheGlyIleProLysValIleGlySerAspAsnGlyProAlaPheValAla 1499  
QY 5085 CAGGTAAGTCAGGACTGGCCAGATATTGGGATTGATTGGAACTGCAATTGTGCATAC 5144  
Db 1500 GlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyr 1519  
QY 5145 AGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAACCACTTAAAGAGACCTTTACT 5204  
Db 1520 ArgProGlnSerSerGlyGlnValGluArgMetAsnLysThrIleLysGluThrLeuThr 1539  
QY 5205 AAATTGACCGCGAGACTGGCGTTAATGATTTGATAGTCTCTCTGCGCTTTGCTGCTTTT 5264  
Db 1540 LysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeuLeuProPheValLeuPhe 1559  
QY 5265 AGGGTTAGGAACACCCCTGGACAGTTTGGCTGACCCCTCTAGAAATTTACTCTACGGGGA 5324  
Db 1560 ArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGlnLeuLeuTyrGlyGly 1579  
QY 5325 CCCCCCACTTGTAGAAATTTGCTCTGTACATAGTGTGCTGCTGCTTTTCCAGCCT 5384  
Db 1580 ProProProLeuValGluIleAlaSerValHisSerAlaAspValLeuLeuSerGlnPro 1599  
QY 5385 TTGTTCTCTAGGCTCAAGGCACTTGAGTGGGTGAGACAAACGAGCTGGAGGCACTCCGG 5444  
Db 1600 LeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArg 1619

Qy	5445	GAGCCTACTCAGGAGGAGAGACTTCGAGATCCCACATCGTTTCCAAAGTGGGAGATTCA	5504
Db	1620	GlulAlaTyrSerGlyGluGlyAspLeuGlnValProHisArgPheGlnValGlyAspSer	1639
Qy	5505	GTCTACGTTAGACCCACCGTCGAGGAAACCTCGAGACTCGTGGGAAGGGCCCTTATCTC	5564
Db	1640	ValTyrValArgArgHisArgAlaSerGlyAsnLeuGluThrArgTyrPylsGlyProTyrLeu	1659
Qy	5565	GTACTCTTTGACCACACCAACGGCTGCAAGTTCGAAGTAATCTCCACTCGATCCATGCA	5624
Db	1660	ValLeuLeuThrThrProThrAlaValLysValGluGlyIleSerThrTrpIleHisAla	1679
Qy	5625	TCCACACGTTAAACCGCGGCCACCTCCCGATTTCGGGGTGGAAAGCCGAAAAAGACTGAA	5684
Db	1680	SerHisValLysLeuAlaProProProAspSerGlyTyrPargAlaGluLysThrGluAsn	1699
Qy	5685	CCCCTTAAGCTTCGCCCTCCATCGCGTGGTTCCTTACTCTGTCATAAACCTCT	5736
Db	1700	ProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsnAsnAsn-SerProGlyGly	1719
Qy	5736	-----LysTyrCysGlySerGluGluSerThrArgGlyGluProLysArgLe	5736
Db	1719	nMetHisProThrLeuSerTrpArgHisLeuProThrArgGlyGluProLysArgLe	1739
Qy	5737	-----uArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrProGly	5738
Db	1739	uArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrProGly	1759
Qy	5739	GACTAAATGGTATGCGCATAGGAGACGACCTGAACTCCCATAAACCTTATCTCTCACCTG	5798
Db	1759	nAlaSerSerLysArgLeuIleAspSerSerAsnProHisArgProLeuSerLeuThrTr	1779
Qy	5799	GTTAAATTAAGTACGCTCCGACAGTAGTATTAATATCAACAACACTCAAGGGAGGCTCTTT	5858
Db	1779	pLeuIleAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProAr	1799
Qy	5859	AGGAACCTGGTGGCTCATCTATACGTTTGCTCTCAGATCAGTTATTCCTAGTCTGACCTC	5918
Db	1799	gGlyThrTrpTrpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSe	1819
Qy	5919	A---CCCCAGATATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACCAACCAATAA	5975
Db	1819	rThrProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGly	1839
Qy	5976	TGGAACAATTCGGGAATCCAGAGATTTCTTTTGTAAACAATGGAACCTGTGTAACTCTC	6035
Db	1839	u---LysTyrCysGlyGlySerGluGluSerPheCysArgArgTrpSerCysValThrSe	1858
Qy	6036	TAATGATGGATATTGGAATGGCCAACTCTCAGCAGGATAGGTAGTTTCTTCTATGT	6095
Db	1858	rAsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVa	1878
Qy	6096	CAACACCTATACCACTCTGCACAAATTAATTAATCTGACCTGGATTAGAACTGGAAGCCC	6155
Db	1878	lAsn-----SerGlyProGlyLysTyrLysValMetLysLeuTyrLysAspLys	1894
Qy	6156	CAAGTGTCTCTCTTCAGACCTTAGATTACTTAAATAAATAGTTTCTACTGAGAAAGGAAACA	6215
Db	1894	sSerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGly	1914
Qy	6216	AGAAAATATCTAAATGGGTAAATGTATGTCTTGGGGAATGGTATAT-----TATGG	6269
Db	1914	nGluAsnIleGlnLysTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVa	1934
Qy	6270	AGGCTCGGGTAAACAACACAGSCTCCATCTTAACCTATTTCGCTCAAAATAAACACAG---	6326
Db	1934	yGlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyTh	1951
Qy	6327	GGAGCCTCCATATAGGACCAAAATACGGTCTTGACGGGTCAAAAGACCCCAACCCCA	6386
Db	1951	rGluProProValAlaValGlyProAspGlyValLeuAlaGluGlnGlyProProAlaLe	1971

QY	6387	AGHACCA-----GG	6390
Db	1971	uGLuProHieAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrGl	1991
QY	6396	ACCATCTCTAAACATAACTTCTGATCAGACCCACCTAGTCT-----AACAGC-----	6444
Db	1991	nProProSerAsnGlyThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGl	2011
QY	6445	----ACGACTAAATGGGGCAAAACCTTTTATAGCCTCATCCAGGGAGCTTTTCAAGCTCT	6500
Db	2011	yValProVallysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle	2031
QY	6501	TAACTCCACGACTCCAGAGGCTACCTCTTCTGTGGCTATGCTTACGTTCTGGGCCCCACC	6560
Db	2031	eAsnSerThAspProAspAlaThrSerSerCysTrpLeuCysLeuSerGlyProPr	2051
QY	6561	TTACTATGAAGGAATGCTAGAGAGGAATTCATGTGACAAAGAAACATAGAGACCA	6620
Db	2051	oTyTyTyGluGlyMetAlaIysGluGlyLysPheAsnValThrLysGluHisArgAsnGl	2071
QY	6621	ATGCACATGGGATCCCAAATTAAGCTTACCTTTACTGAGGTTTCTCGAAAAGGCACCTG	6680
Db	2071	nCysThrTrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCy	2091
QY	6681	CATAGGAAGGTTCCCCATCCCAACACACCTTTGTAAACACACTGAAGCCTTTAATCA	6740
Db	2091	sIleGlyLysAlaProProSerHisGlnHisLeuCysTyTySerThrValValTyTyGluGl	2111
QY	6741	AACCTCTGAGATCAATATCTGTACTCTGTTATGACAGGTGTGGGCATGTAATACTGG	6800
Db	2111	nAlaSerGluAsnGlnTyTyLeuValProGlyTyTyAsnArgTrpTrpAlaCysAsnThrGl	2131
QY	6801	ATTAAACCCCTGTCTTCCACCTTGGTGTTTTAAACCAAACTAAAGATTTTTCATTTATGGT	6860
Db	2131	yLeuThrProCysValSerSerSerValPheAsnGlnSerLysAsnPheCysValMetVa	2151
QY	6861	CCAAATGTTCCCGAGTGTTACTATATCCGAAAAAGCAATCCTTGATGAATATGACTA	6920
Db	2151	iGlnIleValProArgValTyTyHisProGluGluValValLeuAspGluTyTyAspTy	2171
QY	6921	CAGAAATCATCGAANAAGAGAGAACCCATATCTCTGACACTTGTCTGTGATGCTCGACT	6980
Db	2171	rArgTyTyAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLe	2191
QY	6981	TGGAGTGGCAGCAGCTGTAGAAACAGAAACAGCTGCCCTGGTACCGGGACACACAGCAGCT	7040
Db	2191	uGlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLe	2211
QY	7041	AGAAACAGGACTTAGTAACCTACATCGAATGTGTAAACAGAGATCTCCAAGCCCTAGAAAA	7100
Db	2211	uGluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuGluGl	2231
QY	7101	ATCTGTCAGTTAACTGGAGGAATCCCTAACTCTCTTATCTGAAGTAGTCTCTACAGATAG	7160
Db	2231	uSerValSerAsnLeuGluSerLeuThrSerLeuSerGluValValLeuGlnAsnAr	2251
QY	7161	AAGAGGCTTAGATTATTCTTCTAAAGAGGAGGATTATGTCTAGCTTCTGAAGGAGGA	7220
Db	2251	gArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGl	2271
QY	7221	ATGCTGTTTTTATCTGGATCATTCAGGGGCCATCAGAGACTCCATGAACAGCTTAGAGA	7280
Db	2271	uCysCysPheTyTyValAspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGl	2291
QY	7281	AAGTTTGAGAAAGCGTCAAGGGGAAAAGGAAACTATCTCAAGGTGTGTTTGGGGATGGTT	7340
Db	2291	uArgLeuGluArgArgArgGluArgGluArgGluAlaAspGlnGlyTrpPheLysGlyTyTyPh	2311
QY	7341	CAACAGCTCTTTGGTTGGCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	7400
Db	2311	eAsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLe	2331
QY	7401	CTCTCTGTTACTCACAGTGGGCCATGTATTTATTAAACAGTTAATTCCTTTCATTAGAGA	7460

Db 2331 uLeuLeuLeuThrValGlyProCysLeuileAsnArgPheValAlaPheValArgG1 2351  
Qy 7461 ACGAATAAGTGCAGTCCAGATCATGGTACTTAGACAAACAGTACC 7506  
Db 2351 uArgValSerAlaValGlnIleMetValLeuArgGlnTyrGln 2366

RESULT 4

Q8J4V6\_9GAMR PRELIMINARY; PRT; 1720 AA.  
AC Q8J4V6;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Porcine endogenous retrovirus B.  
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=194959;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22172947; PubMed=12185278;  
RA Bartosch B., Weiss R.A., Takeuchi Y.;  
RT "PCR-based cloning and immunocytological titration of infectious  
RT porcine endogenous retrovirus subgroup A and B.";  
RL J. Gen. Virol. 83:2231-2240(2002).  
DR EMBL; AY099324; AAM29194.1; -; Genomic\_DNA.  
DR HSP; P03355; 116J.  
DR SMR; Q8J4V6; 7-98.  
DR MEROPS; A02\_020; -.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR003036; Gag\_p30.  
DR InterPro; IPR000840; G\_retro\_matrix.  
DR InterPro; IPR001995; Peptidase\_A2\_cat.  
DR InterPro; IPR001969; Pept\_Asp\_AS.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF01140; Gag\_MA; 1.  
DR Pfam; PF02093; Gag\_p30; 1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; Rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR Pfam; PF00098; zf-CCHC; 1.  
DR SMART; SM00343; Znf\_C2HC; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS01075; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS00879; RNASE\_H; 1.  
DR PROSITE; PS00158; ZF\_CCHC; 1.  
KW Polyprotein.  
SQ SEQUENCE 1720 AA; 193393 MW; 7F9DCC2F5E834D8A CRC64;

Alignment Scores:

Pred. No.: 0 Length: 1720  
Score: 8821.00 Matches: 1655  
Percent Similarity: 98.4% Conservative: 34  
Best Local Similarity: 96.4% Mismatches: 27  
Query Match: 60.3% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3 (1-8132) x Q8J4V6\_9GAMR (1-1720)

Qy 585 ATGGGACAGACGGTGCAGACCCCTCTTAGTTTGACTCTCGACCATTCGACTGAAGTTAAA 644  
Db 1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisThrThrGluValArg 20  
Qy 645 TCAGGGCTCATATTTGTCAGTTACAGTTAAGAGGGACCTTGGCAGACTTCTGTGTC 704  
Db 21 SerArgAlaHisAsnLeuSerValGlnValLysGlyProThrPheCysAla 40  
Qy 705 TCTGAATGGCCGACATTCGATGTTGGATGCCCATCAGAGGGAGCTTTTAATCTGAGATT 764  
Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
Qy 765 ATCTGCTGCTTTAAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCGATCAGGAG 824  
Db 61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
Qy 825 CCCTATATCTTAGCTGGCAAGATTTGGCAGAGAGATCTCCGCCATCGGTTAAACATGG 884  
Db 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100  
Qy 885 CTGAATAAGCCAAAGACCCAGGTCCCGCAATCTTGGCTCTTGGAGAGAAAACAAACAC 944  
Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
Qy 945 TCGGCTGAAAGAGTCAAGCCCTCTCTCATATATACCCCGAGATTGAGGAGCCACCCGGCT 1004  
Db 121 SerAlaGluLysValGluProSerProArgIleTrpProGluIleGluGluProThr 140  
Qy 1005 TGGCCGGAACCCCAATCTGTTCCCGACCCCTTATTCGCACAGGTGGCGGAGGGA 1064  
Db 141 TrpProGluProGlnProValProProProTrpProAlaGlnGlyAlaValArgGly 160  
Qy 1065 CCCTTTGCCCTCTCGAGCTCCGGGCTGGAGGAGCTCTCGAGGAGTCTCGAGGCGG 1124  
Db 161 ProSerAlaProProGlyAlaProValGluGlyProAlaAlaGlyThrArgSerArg 180  
Qy 1125 AGGGGCGCCACCCGGAGCGGACAGAGATCGCAGATTCACCGCTTGGCCATCGTACGGC 1184  
Db 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly 200  
Qy 1185 CTCCACACACCGGGGGCCAAATTCAGCCCTCTCAGTATTCAGTATTCCTCTGAGAT 1244  
Db 201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTrpProPheSerSerAlaAsp 220  
Qy 1245 CTCTATAATTGGAAACTAACCATCCCTCTCGAGGATCCCAACCCCTCAGCGG 1304  
Db 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
Qy 1305 TTGGTGGAGTCCCTTATGTTCTCTCACAGCCCTTCTGGATGATTTGTCAACAGCTCTG 1364  
Db 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnLeuLeu 260  
Qy 1365 CAGACACTTTCACAAACCGAGGAGCGAGAGAAATTCATTAGAGGCTAGAAAAATGTT 1424  
Db 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280  
Qy 1425 CCTGGGGCGAGCGGGCGACCCCGGTTCGAAATGAGATTGACATGGGATTCCTCCTTA 1484  
Db 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
Qy 1485 ACTCGCCCGGTTGGGACTACAAACAGCGCTGAAAGGTAGGAGAGCTTGAATAATCTATCGC 1544  
Db 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
Qy 1545 CAGGCTCTGGTGGCGGGTCTCCGGGGGCGCTCAAGACGGCCCACTAATTTGGCTAGGTA 1604  
Db 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
Qy 1605 AGAGNAGTGAATCAGGAGCGGATGAACCCCTCTGTTTCTTCTGAGAGGCTCTTGAA 1664  
Db 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360



QY 1665 GCCTTCAGGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAGCCTCAGTCGCT 1724  
Db |||||:::|||||  
361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
QY 1725 TTGCGCCTTTATAGACACAGTCAGCCCTTGATATTAGAAAAGAGCTTCAGAGACTCGAAGGG 1784  
Db |||||:::|||||  
381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
QY 1785 TTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAGCGAGAGAAAGTATATTACAAAGG 1844  
Db |||||:::|||||  
401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg 420  
QY 1845 GAGACAGAGAAGAAAGGGAACAAGAAAGAGAGAGAAAGAGAGAGAAAGGAGGAAGA 1904  
Db |||||:::|||||  
421 GluThrGluGluGlyLysGluGlnArgLysGluArgGluGluArgGluGluArg 440  
QY 1905 CGTAAATAAGCGCAAGAGAGAAGAAATTGACTAAGATCTTGCGTCAGTGGTTGAAAGGAAA 1964  
Db |||||:::|||||  
441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValValGluGlyLys 460  
QY 1965 AGCAATACGGAAGACAGAGAGAGATTTTAGGNAATTTAGCTCAGGCCCTAGACAGTCAGGG 2024  
Db |||||:::|||||  
461 SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
QY 2025 AACCTGGGCAATAGGACCCCACTCGACAAGGACCAATGTGCATATTGTAAAGAAAGAGGA 2084  
Db |||||:::|||||  
481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500  
QY 2085 CACTGGGCAGGAARCTGCCCCAAGAGGAAACAAAGGACCAAGAGCTCTAGCTCTAGAA 2144  
Db |||||:::|||||  
501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520  
QY 2145 GAAGATAAAGATTAGGGGAGACGGGGTTTCGAGACCCCTCCCGAGCCCGAGGTTAACTTTG 2204  
Db |||||:::|||||  
521 GluAspLysAsp\*\*GlyArgGlySerAspProLeuProGluProArgValThrLeu 540  
QY 2205 AAGGTGAGGGGCAACAGATTGAGTTCCTGGTTCATACCGGAGCGAAACATTCAGTGCTA 2264  
Db |||||:::|||||  
541 LysValGluGlyGlnProValGluPheLeuValAspThrGlyAlaGluHisSerValLeu 560  
QY 2265 CTACAGGCATTAGGAAAATCTAAAGATAAAATCTCTGGGTGATGGGTGCCACAGGGCAA 2324  
Db |||||:::|||||  
561 LeuGlnProLeuGlyLysLeuLysGluLysLysSerTrpValMetGlyAlaThrGlyGln 580  
QY 2325 CAACAGTATCCATCGACTACCCGAAGAACAGTTCAGCTTGGAGTGGGACGGGTAAACCCAC 2384  
Db |||||:::|||||  
581 ArgGlnTyrProTrpThrThrArgArgThrValAspLeuGlyValGlyArgValThrHis 600  
QY 2385 TCGTTTCTGGTTCATACCTGAGTGCACACCCCTCTTAGGTAGAGACTTATTACCAACAG 2444  
Db |||||:::|||||  
601 SerPheLeuValIleProGluCysProValProLeuLeuGlyArgAspLeuLeuThrLys 620  
QY 2445 ATGGGAGCACAATTTCTTTTGAACAAGGGAACCCAGAAGTGTCTGCAAAATAACAAACCT 2504  
Db |||||:::|||||  
621 MetGlyAlaGlnIleSerPheGluGlnGlyArgProGluValSerValAsnAsnLysPro 640  
QY 2505 ATCACTGTGTGACCTCCCAATTAGATGACGAATATCGACTATACTCTCCCTAGTAAAG 2564  
Db |||||:::|||||  
641 IleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeuTyrSerProGlnValLys 660  
QY 2565 CCTGATCAAAATATACAAATTCGTGTGGAACAGTTTCCCAAGCTCGGGCAGAAACCGCA 2624  
Db |||||:::|||||  
661 ProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGlnAlaTrpAlaGluThrAla 680  
QY 2625 GGGATGGGTTGGCAAGCAAGTTCCCCCAACAAGTTATTCAACTGAAGGGCCAGTGCACCA 2684  
Db |||||:::|||||  
681 GlyMetGlyLeuAlaLysGlnValProProGlnValIleGlnLeuLysAlaSerAlaThr 700  
QY 2685 CCAGTGTCAAGTACAGTACCCCTTGATGATAAGAGCTCAAGAGGAATTCGGCCCGCAT 2744  
Db |||||:::|||||  
701 ProValSerValArgGlnTyrProLeuSerArgGluAlaArgGluGlyIleTrpProHis 720  
QY 2745 GTCCAAAGATTAACTCAACAGGGCATCTAGTTCCTGTGCCAATCTCCTCGGAATACTCCC 2804

Db |||||:::|||||  
721 ValGlnArgLeuIleGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrPro 740  
QY 2805 CTGCTACCGGTTAGAAAGCCTGGGACTTAATGACTATCGACAGCTACAGGACTTGAGAGAG 2864  
Db |||||:::|||||  
741 LeuLeuProValArgLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGlu 760  
QY 2865 GTCAATAAACGGGTGCGAGGATATACACCCCAACAGTCCCGAACCTTATACCTCTCTGTG 2924  
Db |||||:::|||||  
761 ValAsnLysArgValGlnAspIleHisProThrValProAsnProTyrAsnLeuLeuSer 780  
QY 2925 GCTCTCCACCCCAACGAGCTGCTATACAGTATTGAGCTTAAAGAGTGCCTCTCTTCGC 2984  
Db |||||:::|||||  
781 AlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeuLysAspAlaPhePheCys 800  
QY 2985 CTGAGATTACACCCCACTAGCCAAACACTTTTTCCTTCGATCGAGAGATCCAGGTACG 3044  
Db |||||:::|||||  
801 LeuArgLeuHisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyThr 820  
QY 3045 GGAAGAACCGGCGAGCTCACCTGGACCCGACTGCCCAAGGGTTCAAGAACTCCCCGACC 3104  
Db |||||:::|||||  
821 GlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThr 840  
QY 3105 ATCTTTGACGAAGCCCTACACAGAGACTCTGGCCCAACTTCAGGATCCAAACCCCTCAGGTG 3164  
Db |||||:::|||||  
841 IlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnVal 860  
QY 3165 ACCCTCTCCAGTACGTGGATGACCTGCTCTTCGGGGAGCCACCACCAAGAGGACTGCTTA 3224  
Db |||||:::|||||  
861 ThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAlaThrLysGlnAspCysLeu 880  
QY 3225 GAAGGCACGAAGCACACTACTGCTGGATTCCTGACCTAGGCTACAGAGCTCTGCTAAG 3284  
Db |||||:::|||||  
881 GluGlyThrLysAlaLeuLeuLeuLeuLeuSerAspLeuGlyTyrArgAlaSerAlaLys 900  
QY 3285 AAGCCCCAGATTTGACGAGAGAGGTAAACATCTTGGGGTACAGTTTTCGCGGACGGGCAG 3344  
Db |||||:::|||||  
901 LysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyrSerLeuArgGlyGln 920  
QY 3345 CGATGCTGACGAGGCGACCGGAAGAAACCTGTAGTCAGATACCGGGCCCAACACAGCC 3404  
Db |||||:::|||||  
921 ArgTrpLeuThrGluAlaArgLysLysThrValValGlnIleProAlaProThrAla 940  
QY 3405 AAACAAATGAGAGAGTTTTCGGGACAGCTGGATTTTGACAGCTGTGAGTCCCGGGTTT 3464  
Db |||||:::|||||  
941 LysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPhe 960  
QY 3465 GCAGACCTTAGCAGCCCCACTCTACCCCTAAACCAAGAAAGGGGAATTCCTCTGGGCT 3524  
Db |||||:::|||||  
961 AlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLysGlyGlyPheSerTrpAla 980  
QY 3525 CTTGAGCACAGAGGAGGATTTGATGCTATCAAAAGCCCTGCTGAGCGGCACCTGCTCTG 3584  
Db |||||:::|||||  
981 ProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuLeuSerAlaProAlaLeu 1000  
QY 3585 GCCCTCCCTCAGCTAAACCTTTTACCTTTATGTGGATGAGCGCTAAGGGAGTAGTACC 3644  
Db |||||:::|||||  
1001 AlaLeuProAspValThrLysProPheThrLeuTyrValAspGluArgLysGlyValAla 1020  
QY 3645 CGGGAGCTTTTAAACCCAAACCTTAGGACCATGGAGAAAGACTGTGCGCTACCTCTCAAG 3704  
Db |||||:::|||||  
1021 ArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLys 1040  
QY 3705 AAGCTCATCTCTAGCCAGTGGTTGGCCCATATGCTCCCTGAAGGCTATCGCAGCTGTGGCC 3764  
Db |||||:::|||||  
1041 LysLeuAspProValAlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaValAla 1060  
QY 3765 ATACTGCTCAAGGACGCTGACAAATTCACCTTTGGGACAGAAATATAACTGTAAATAGCCCC 3824  
Db |||||:::|||||  
1061 IleLeuValLysAspAlaAspLysLeuThrLeuGlnAsnIleThrValIleAlaPro 1080  
QY 3825 CATGCATTTGGAGAACATCTGTTCCGACGCCCCACGACCGATGGATGACCAACGCCCGCATG 3884



Db 1081 HisAlaLeuGluAsnIleValArgGlnProProAspArgTrpMetThrAsnAlaArgMet 1100  
QY 3885 ACCACATATCAAGCCCTGCTTCTCACAGAGAGGTACGTTTCGCTCCACAGCCGCTCTC 3944  
Db 1101 ThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPheAlaProProAlaAlaLeu 1120  
QY 3945 AACCTGCCACTCTCTCCCTGAAGAGACTGATGAACAGTGACTCATGATTGCCATCAA 4004  
Db 1121 AsnProAlaThrLeuLeuLeuProGluGluThrAspGluProValThrHisAspCysHisGln 1140  
QY 4005 CTATTGATTGAGAGACTGGGCTCCGCAAGGACCTTACAGACATACCCGCTGACTGGAGAA 4064  
Db 1141 LeuLeuIleGluThrGlyValArgLysAspLeuThrAspIleProLeuThrGlyGlu 1160  
QY 4065 GTGCTAACCTGGTTCACTACGAGCAGCTATGCTGGAAGTAAAGAGATGGCTGGG 4124  
Db 1161 ValLeuThrTrpPheThrAspGlySerSerTyrValValGluGlyLysArgMetAlaGly 1180  
QY 4125 GCGCGGTGGTGGACGGACCCGACCATCTGGGCGAGCCTGCGCGGAGGAACTTCA 4184  
Db 1181 AlaAlaValValAspGlyThrArgThrIleTrpAlaSerSerLeuProGluGlyThrSer 1200  
QY 4185 GCACAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGGGCTGGCGGAGGGAATCC 4244  
Db 1201 AlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSer 1220  
QY 4245 ATAACATTTATACGGACAGCAGTATGCTTTGCGACTGCACACGTACATGGGGCCATC 4304  
Db 1221 IleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIle 1240  
QY 4305 TATAACAAGGGGTGCTTACTCTGACGAGGGAGGAATAAAGAACAAAGAGGAAATT 4364  
Db 1241 TyrLysGlnArgGlyLeuLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluIle 1260  
QY 4365 CTAAGCCTATTAGAAGCGTACATTTACCAAAAGGCTAGCTATTATACACTGCTCTGGA 4424  
Db 1261 LeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIleIleHisCysProGly 1280  
QY 4425 CATCAGAAAGCTTAAAGATCTCATATCCAGAGGAAACAGATGGCTGACCCGGTTCGCAAG 4484  
Db 1281 HisGlnLysAlaLysAspLeuIleSerArgLysGlnMetAlaAspArgValAlaLys 1300  
QY 4485 CAGCGACCCAGGGTGTAACTTCTGCTTATATAGAAATGCCCAGGCCAGAACCC 4544  
Db 1301 GlnAlaAlaGlnAlaValAsnLeuLeuProIleIleGluThrProLysAlaProGluPro 1320  
QY 4545 AGACGACAGTACACCTTAGAAGACTGCGAAGAGATAAAGATAGACAGTCTCTGAG 4604  
Db 1321 ArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGlu 1340  
QY 4605 ACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAGGAATCTCGCCCAACAAAGAGGG 4664  
Db 1341 ThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGly 1360  
QY 4665 TTAGAATATGTCACAGATACATGCTCTAACCCACTAGGAACCTAAACACCTGCGAGCAG 4724  
Db 1361 LeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGln 1380  
QY 4725 TTGCTCAGACATCCCTTATCATGTTCTGAGGCTACCGAGTGGCTGACTCGGTGTC 4784  
Db 1381 LeuValArgThrSerProTyrHisValLeuArgLeuProGlyValAlaAspSerValVal 1400  
QY 4785 AAACATTGTGTGCGCTGCCAGCTGGTGTAAATGCTTCCAGAAATCTCTCCAGAAATGCTCCAGGAG 4844  
Db 1401 LysHisCysValProCysGlnLeuValAsnAlaAsnProSerArgIleProProGlyLys 1420  
QY 4845 AGACTAAGGGAGGCCACCCAGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCG 4904  
Db 1421 ArgLeuArgGlySerHisProGlyAlaHisTrpGluValAspPheThrGluValLysPro 1440  
QY 4905 GCTAAATACGGAACAAATACCTATTGGTTTTGTAGACACCTTTTCAGATGGGTAGAG 4964  
Db 1441 AlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpValGlu 1460

QY 4965 GCTTATCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAAAAATACTGGAAGAAATT 5024  
Db 1461 AlaTyrProThrLysLysGluThrSerThrValValAlaLysLysIleLeuGluGluIle 1480  
QY 5025 TTTCCAAGATTGGAAATACCTAAGGTAAATAGGGTCAGACAATGTGTCCAGCTTTGTGGCC 5084  
Db 1481 PheProArgPheGlyIleProLysValIleGlySerAspAsnGlyProAlaPheValAla 1500  
QY 5085 CAGTTAAGTCAGGAGCTGGCAAGATATTGGGATTCATTGGAAATGCTATTGTGTCATAC 5144  
Db 1501 GlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyr 1520  
QY 5145 AGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAACCAATTAAGAGAGACCTTACT 5204  
Db 1521 ArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThr 1540  
QY 5205 AAATTGACCGCGAGACTGGCGTTAATGATTGATAGTCTCTCTGCGCTTTGTGCTTTT 5264  
Db 1541 LysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeuLeuProPheValLeuPhe 1560  
QY 5265 AGGTTAGGAACACCCCTGGACAGTGTGGGCTGACCCCTATGAATTTACTCTACGGGGA 5324  
Db 1561 ArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGluLeuLeuTyrGlyGly 1580  
QY 5325 CCCCCCCTAGTTGGTAGAAATTGCTTCTGTACATAGTGTCTGACGTGCTGCTTTCCAGCCT 5384  
Db 1581 ProProProLeuValGluIleAlaSerValHisSerAlaAspValLeuLeuSerGlnPro 1600  
QY 5385 TTGTTCTTAGGCTCAAGGCACCTGTAGTGGTGAGACCAACAGAGCTGGAGGCACTCCGG 5444  
Db 1601 LeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArg 1620  
QY 5445 GAGGCTACTCAGGAGGAGAGACTTGCAGATCCCATCTGTTCCCAAGTGGGAGATTCA 5504  
Db 1621 GluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArgPheGlnValGlyAspSer 1640  
QY 5505 GTCTAGCTTAGAGCCACCGTGCAGGAAACCTCGAGACTCGTGGAGAGGCGCTTATCTC 5564  
Db 1641 ValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProTyrLeu 1660  
QY 5565 GTACTTTGACCAACCAACCGCTGTGAAGTGAAGGAATCTCCACTGGATCCATGCA 5624  
Db 1661 ValLeuLeuThrThrProThrAlaValLysValGluGlyIleSerThrTrpIleHisAla 1680  
QY 5625 TCCACGTTAAACCGCGCCACCTCCGATTCGGGTGGAAAGCCGAAAGAGACTGAAAT 5684  
Db 1681 SerHisValLysLeuAlaProProAspSerGlyTrpArgAlaGluLysThrGluAsn 1700  
QY 5685 CCCCTTAAGCTTCGCTCCATCGCTGCTTCTTACTCTGCTCAATAAC 5732  
Db 1701 ProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsnAsnAsn 1716

## RESULT 5

Q8J4V8\_9GAMR  
ID Q8J4V8\_9GAMR PRELIMINARY; PRT; 1718 AA.  
AC Q8J4V8;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Porcine endogenous retrovirus A.  
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=194958;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22172947; PubMed=12185278;  
RA Bartosch B., Weiss R.A., Takeuchi Y.;  
RT "PCR-based cloning and immunocytological titration of infectious  
porcine endogenous retrovirus subgroup A and B."  
RL J. Gen. Virol. 83:2231-2240(2002).  
DR EMBL; AY099323; AAM29192.1; -; Genomic\_DNA.

DR HSSP; P03355; I16J.  
 DR SMB; O8JAV8; 7-98.  
 DR MEROPS; A02.020; -.  
 DR GO; GO:0019028; C: viral capsid; IEA.  
 DR GO; GO:0004190; F: aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F: DNA binding; IEA.  
 DR GO; GO:0004523; F: ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F: RNA binding; IEA.  
 DR GO; GO:0003964; F: RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0005198; F: structural molecule activity; IEA.  
 DR GO; GO:0006310; P: DNA recombination; IEA.  
 DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P: RNA-dependent DNA replication; IEA.  
 DR GO; GO:0019068; P: viral assembly; IEA.  
 DR InterPro; IPR003036; Gag\_p30.  
 DR InterPro; IPR000840; G\_retro.matrix.  
 DR InterPro; IPR001995; Peptidase A2 cat.  
 DR InterPro; IPR001969; Pept\_Asp\_AS\_  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVtse.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF011140; Gag\_MA; 1.  
 DR Pfam; PF02093; Gag\_p30; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; Rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR Pfam; PF00098; zf-CCHC; 1.  
 DR SMART; SM00343; Znf\_C2HC; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS50879; RNASE\_H; 1.  
 DR PROSITE; PS0158; ZF\_CCHC; 1.  
 DR Polyprotein.  
 KW POLYPEPTIDE.  
 SQ SEQUENCE 1718 AA; 193115 MW; 4E86215286C59246 CRC64;

Alignment Scores:  
 Pred. No.: 0 Length: 1718  
 Score: 8815.50 Matches: 1655  
 Percent Similarity: 98.3% Conservative: 33  
 Best Local Similarity: 96.3% Mismatches: 29  
 Query Match: 60.2% Indels: 1  
 DB: 2 Gaps: 1

US-10-723-552-3 (1-8132) x Q8JAV8\_9GMR (1-1718)

QY 585 ATGGAGCAGACGCTGACGACCCCTCTTAGTTGACTCTCGACCATTCGACTGAAGTTAA 644  
 DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20  
 QY 645 TCCAGGGCTCATAATTTGTCAAGTTTCAGTTTAAAGAGGACCTTCGCGACATTCCTGTGTC 704  
 DB 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysAla 40  
 QY 705 TCTGAATGGCCGACATTCGATGTTGGATGGCCATCAGAGGGACCTTAATTCGAGATT 764  
 DB 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
 QY 765 ATCTGGCTGTTAAAGCAGATTATTTTTCAGACTCGACCCGGCTCTCATCCCGATCAGAG 824  
 DB 61 IleuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
 QY 825 CCCTATATCCCTGCGCAAGATTTCGACAGGATCTCCGCCATGGTGTAAACCATGG 884  
 DB 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100  
 QY 885 CTGATAAGCCAGAACCGAGGCTCCCGAATTCGGCTCTTCGGAGGAAACAAACAC 944  
 DB 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
 QY 945 TCCGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGACCCCGCT 1004  
 DB 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500

DB 121 SerAlaGluLysValGluProSerProArgIleTyrProGluIleGluGluProProThr 140  
 QY 1005 TGGCCGGAACCCCAATCTGTTCCCCACCCCTATCTGCGACAGGGTCCGCGAGGGA 1064  
 DB 141 TrpProGluProGlnProValProProProProProProAlaGlnGlyAlaValArgGly 160  
 QY 1065 CCCTTTGCCCTCTCTGGAGCTCGGCGGTGGAGGACCTGCTGCAGGACTCGGAGCGG 1124  
 DB 161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180  
 QY 1125 AGGGGCGCCACCCCGAGCGGACAGATCCGACATTCGCGTCCGCGACATCAGGC 1184  
 DB 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly 200  
 QY 1185 CTCCCCACACGGGGGCCAAATTGCAGCCCTCCAGATTTCGGCCCTTTCTTCGCGAT 1244  
 DB 201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
 QY 1245 CTCTATAATTGGAAACTAACCATCCCTTCTCGGAGGATCCCAACGCTCACGGG 1304  
 DB 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
 QY 1305 TTGGTGGAGTCCCTTATGTTCTCTCACGACCTACTTTGGGATGATTCACACGCTCTG 1364  
 DB 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeu 260  
 QY 1365 CAGACACTCTTCACACCGGAGGAGGAGAGATTCATTATAGAGGCTAGAAAAATGTT 1424  
 DB 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280  
 QY 1425 CCTGGGGCCGACGGGACCCACGCGGTTCGAAATGAGATTCATGGGATTTCCCTTA 1484  
 DB 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluLeuAspMetGlyPheProLeu 300  
 QY 1485 ACTCGCCCGGTTGGGACTACAACACGCTGAAGTTCAGGAGAGCTTGAATAATCTATCGC 1544  
 DB 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
 QY 1545 CAGGCTCTGTGGGGGTCTCCGGGGGCTCAAGACGGCCCACTAATTTGGCTAAGGTA 1604  
 DB 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
 QY 1605 AGAGNAGTGTATGACGAGGACCGAATGAACCCCTCTCTGTTTCTTGAGAGCTCTTGAA 1664  
 DB 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
 QY 1665 GCCTTCAGGCGGTACACCCCTTTTGATCCCACTCAGAGGCCCAAAAGCCTCAGTGGCT 1724  
 DB 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
 QY 1725 TTGGCCTTTATAGGACAGTTCAGCCTTCGATATTAGAAGAAGCTTCAGAGCTGGAAGG 1784  
 DB 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
 QY 1785 TTACAGGAGGCTGAGTTACGTGATCTAGTAGGAGGACAGAAAGTATATTACAAAGG 1844  
 DB 401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg 420  
 QY 1845 GAGACAGAAGAAGGAAACAAAGAAAGAGAGAGAAAGAGAGAAAGAGAGAGAGAGAG 1904  
 DB 421 GluThrGluGluLysGluGlnArgLysGluLysGluArgGluGluArgGluGluArg 440  
 QY 1905 CGTAATAACCGCAGAGAGAAGATTTGACTAAGATCTTGGCTGCGAGTGGTGAAGGAAA 1964  
 DB 441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys 460  
 QY 1965 AGCAATACGAAAGAGAGAGAGATTTTAGGAAATTAGGTCAGGCCCTTAGACATCAGGG 2024  
 DB 461 SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
 QY 2025 AACCTGGCGCAATAGGACCCCACTCGACAAGACCAATGTGCAATTTGTAAAGAAGAGGA 2084  
 DB 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500



Db 1221 ILeasnlleYrThrApsSerArgTyRAlaPhealThrAlaHisValHisGlyAlaIle 1240  
Qy 4305 TATAAACAAGGGGTGCTTACTCTCAGCAGGGAGGGAATAAAGACAAGAGAAATT 4364  
Db 1241 TyRlysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluIle 1260  
Qy 4365 CTAGACCTATTAGAACGGTCATTTACCAAAAAGGCTAGCTATTATACATGTCCTCGGA 4424  
Db 1261 LeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIleIleHisCysProGly 1280  
Qy 4425 CATCAGAAAGTAAAGATCTCATATCCAGAGGAACACAGATGGCTGACCGGGTTGCCAAG 4484  
Db 1281 HisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMetAlaAspArgValAlaLys 1300  
Qy 4485 CAGCAGCCAGGGTGTAACTCTCTGCTATATAGAAATGCCAAAGCCCCAGAACCC 4544  
Db 1301 GlnAlaGlnAlaValAsnLeuLeuProIleIleGluThrProLysAlaProGluPro 1320  
Qy 4545 AGACGACAGTACACCCTAGAGACTGGCAAGATATAAAAGATAGACCACTCTCTCAG 4604  
Db 1321 ArgArgGlnTyRThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGlu 1340  
Qy 4605 ACTCCGAAGGCACTGCTATACCTCAGATGGGAAGGAAATCTCCGCCCAACAAGAGGG 4664  
Db 1341 ThrProGluGlyThrCysTyRThrSerTyRlysGluIleLeuProHisLysGluGly 1360  
Qy 4665 TTAGAATATGTCACACAGATACATCGTCTAACCCACCTAGGAATAAACACCTCGAGAG 4724  
Db 1361 LeuGluTyRValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGln 1380  
Qy 4725 TTGCTCAGAAATCCCTTATCATGTTCTGAGGCTACCAAGAGTGGCTGACTCGGTGTC 4784  
Db 1381 LeuValArgThrSerProTyRHisValLeuArgLeuProGlyValAlaAspSerValVal 1400  
Qy 4785 AAACATTGTGCTGCCTGCACCTGTTTAATCTTAATCTTCCAGAATGCCCTCCAGGGAAG 4844  
Db 1401 LysHisCysValProCysGlnLeuValAsnAlaAsnProSerArgIleProProGlyLys 1420  
Qy 4845 AGACTAAGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTAGAGTTAAAGCCG 4904  
Db 1421 ArgLeuArgGlySerHisProGlyAlaHisTrpGluValAspPheThrGluValLysPro 1440  
Qy 4905 GCTAAATACGGAACAATACCTATGCTGTTTCTAGACACCTTTTCCAGGATGGTAGAG 4964  
Db 1441 AlalysTyRGlYAsnLysTyRLeuLeuValPheValAspThrPheSerGlyTrpValGlu 1460  
Qy 4965 GCTTATCTCTAAGAAAGAGACTTCAACCGTGGTGGCTAAATAAAATACTCGAAGAAATT 5024  
Db 1461 AlaTyRProThrLysLysGluThrSerThrValValAlaLysLysIleLeuGluGluIle 1480  
Qy 5025 TTTCCAGATTGGAATACCTAAGTAATAGGTCAGCAATAGTCCAGCTTTTGTGGCC 5084  
Db 1481 PheProArgPheGlyIleProLysValIleGlySerAspAsnGlyProAlaPheValAla 1500  
Qy 5085 CAGGTAAGTACAGGACTGCGCAAGATATTGGGGATTGATTGGAACCTGATTGTCATAC 5144  
Db 1501 GlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyR 1520  
Qy 5145 AGACCCAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATTAAGACACCTTACT 5204  
Db 1521 ArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThr 1540  
Qy 5205 AAATGACCGCGAGACTGGCGTTAATGATTGGATAGCTCTCTGCGCTTTTGTGCTTTT 5264  
Db 1541 LysLeuThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeuPhe 1560  
Qy 5265 AGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCTATGATTAATTAATCTAGCGGGA 5324  
Db 1561 ArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyRLeuLeuLeuTyR-GlyGly 1580  
Qy 5325 CCCCCCATTTGGTAGAAATGCTTCTGTACATAGTGTGACGTGCTCTTCCAGCCT 5384  
|||||

Db 1581 ProProProLeuAlaGluIleAlaPheAlaHisSerAlaAspValLeuLeuSerGlnPro 1600  
Qy 5385 TTGTTCTCTAGGCTCAGGCACCTGAGTGGTGGAGACAACGAGCGTGGAGCACTCCGG 5444  
Db 1601 LeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArg 1620  
Qy 5445 GAGSCCTACTCAGGAGGAGGAGACTTCAGATCCACATCGTTTCCAAAGTGGGAGATTCA 5504  
Db 1621 GluAlaTyRSer---GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSer 1639  
Qy 5505 GTCTACCTTAGACCCACCTGCGAGGAACTCGAGACTCGGTGGAAAGGGCCCTTATCTC 5564  
Db 1640 ValTyRValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProTyRLeu 1659  
Qy 5565 GTACTTTTGACCAACACGCGCTGCAAGTCCGAAGGAATCTCCACCTCGATCCATGCA 5624  
Db 1660 ValLeuLeuThrThrProThrAlaValLysValGluGlyIleProThrTrpIleHisAla 1679  
Qy 5625 TCCACACGTTAAACCGCGCCACCTCCCGATTTCGGGGTGGAAAGCCGAAAGACACTGAAAT 5684  
Db 1680 SerHisValLysProAlaProProAspSerGlyTrpLysAlaGluLysThrGluAsn 1699  
Qy 5685 CCCTTAAGCTTCGCTCCATCGCGTTCCTTACTCTGTCAATAACCTCTCA 5738  
Db 1700 ProLeuLysLeuArgLeuHisArgValValProTyRSerValAsnAsnSerSer 1717  
RESULT 6  
Q5Q08\_9GAMR  
ID Q5Q08\_9GAMR PRELIMINARY; PRT; 1193 AA.  
AC Q5Q08;  
DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
DE Pol protein (fragment).  
OS Porcine endogenous retrovirus C/A.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gamma retrovirus; 1-Mammalian type C virus group.  
OX NCBI\_TaxID=286072;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=A14/220;  
RX PubMed=15564496; DOI=10.1128/JVI.78.24.13880-13890.2004;  
RA Bartosch B., Stefanidis D., Myers R., Weiss R., Patience C.,  
Takeuchi Y.;  
RT "Evidence and consequence of porcine endogenous retrovirus  
recombination.";  
RL J. Virol. 78:13880-13890 (2004).  
CC -I- SIMILARITY: Contains 1 reverse transcriptase domain.  
DR EMBL; AY570980; AAT7167.1; -; Genomic DNA.  
DR GO; GO:0004190; P:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; P:DNA binding; IEA.  
DR GO; GO:0008233; P:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR009007; Pept\_Aspartc\_cat.  
DR InterPro; IPR001995; Peptidase\_A2\_cat.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR012337; RNaseH\_fold.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVtse.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT; 1; 1.  
DR PROSITE; PS50175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS50994; INTEGRASE; 1.  
DR PROSITE; PS50879; RNASE\_H; 1.  
DR PROSITE; PS50878; RT\_POL; 1.  
KW Aspartyl protease; Protease.

```
FT  NON_TER      1 1
SQ  SEQUENCE    1193 AA; 133310 MW;  9F8EC2AD1A6CE10  CRC64;

Alignment Scores:
Pred. No.:      0      Length:      1193
Score:          6248.50  Matches:      1187
Percent Similarity: 99.7%  Conservative:  2
Best Local Similarity: 99.5% Mismatches:    3
Query Match:      42.7%  Indels:      1
DB:              2      Gaps:      1

US-10-723-552-3 (1-8132) x Q5QG08_9GMR (1-1193)

QY  2160 GGGAGACGGGTTCCGACCCCTCCCGAGCCCGAGGTAACTTTGAAGGTGGAGGGCAA 2219
DB  1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysValGluGlyGln 20

QY  2220 CCAGTTCAGTTCCTGGTGTATACCGGAGCGAACAATTCAGTGTCTACTACAGCCATTAGGA 2279
DB  21 ProValGluPheLeuValAspThrGlyAlaLysHisSerValLeuLeuGlnProLeuGly 40

QY  2280 AAACTAAAGATAAAAAATCTCGGTGATGGTGCCACAGGGCAACACAGTATCCATGG 2339
DB  41 LysLeuLysAspLysLysSerTrpValMetGlyAlaThrGlyGlnGlnGlnTyrSerTrp 60

QY  2340 ACTACCCGAAGAACAGTTGACTTGGGAGTGGACGGGTAAACCCCACTCGTTTCTGGTCATA 2399
DB  61 ThrThrArgArgThrValAspLeuGlyValGlyArgValThrHisSerPheLeuValIle 80

QY  2400 CCTGAGTGGCCAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATGGAGGACAAATT 2459
DB  81 ProGluCysProAlaProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100

QY  2460 TCTTTTCAACAGGGAACACAGAGTCTCTGCAATAACAAACCTATCATCTGTGTGACCC 2519
DB  101 SerPheGluGlnGlyLysProGluValSerAlaAsnAsnLysProIleThrValLeuThr 120

QY  2520 CTCCAATTAGATGACGAATATCGACTACTCTCCCTAGTAAAGCCCTGATCAAAATATA 2579
DB  121 LeuGlnLeuAspAspGluTyrArgLeuTyrSerProLeuValLysProAspGlnAsnIle 140

QY  2580 CAATTCTGGTTGGAAACAGTTTCCCAAGCTGGGAGCAAAACCGAGGGATGGTTGGCA 2639
DB  141 GlnPheTrpLeuGluGlnPheProGlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAla 160

QY  2640 AAGCAAGTTCCTCCCAACAGTTATTCAACTGAAGCCAGTCCACACCAAGTGTCAATCAGA 2699
DB  161 LysGlnValProProGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArg 180

QY  2700 CAGTACCCCTTGATTAAGAAAGCTCAAGAGGAATTCGGCCGCATGTCCAAAGATTAAATC 2759
DB  181 GlnTyrProLeuSerLysGluAlaGlnGluGlyIleArgProHisValGlnArgLeuIle 200

QY  2760 CAACAGGGCATCTAGTCTCTGTCCTCAATCTCCCTGGAACTACTCCCTGCTACCGGTAGA 2819
DB  201 GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLeuProValArg 220

QY  2820 AAGCCTGGGACTAATGACTATCGACAGTACAGAGCTTGAGAGAGGTCAATAAACGGTG 2879
DB  221 LysProGlyThrAsnAspTyrArgProValGlnAspLeuArgLysValAsnLysArgVal 240

QY  2880 CAGGATATACACCCCAACAGTCCCGAACCTTTAATACCTCTGTGTGCTCTCCACCCCAA 2939
DB  241 GlnAspIleHisProThrValProAsnProTyrAsnLeuLeuCysAlaLeuProProGln 260

QY  2940 CGAGCTGGTATACAGTATTGGACTTAAAGATGCCCTTCTTCTGCTCCGTGAGATTACCC 2999
DB  261 ArgSerTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisPro 280

QY  3000 ACTAGCCCAACCTTTTTCCTTCGAATGGAGATCCAGGTACGGGAGAAACCGGGCAG 3059
DB  281 ThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln 300
```

```
QY  3060 CTCACCTGGACCCCGACTGCCCCCAAGGGTTCAAGAACTCCCGACCATCTTTTGACGAAGCC 3119
DB  LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePheAspGluAla 320

QY  3120 CTACACAGACACCTGGCCAACTTCAGGATCCAACACCTCAGGTGACCTCTCTCAGTAC 3179
DB  LeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnTyr 340

QY  3180 GTGATGACCTGCTTCTGGCGGAGCCACCAACAGGACTGCTTAGAAGGACGACCAAGGCA 3239
DB  ValAspAspLeuLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGluGlyThrLysAla 360

QY  3240 CTACTGCTGAATGTCTGACCTAGGCTACAGAGCTCTGCTAAGAAGGCCACAGATTTC 3299
DB  LeuLeuLeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCys 380

QY  3300 AGGAGAGAGTAACATACTTGGGGTACAGTTTGGGGACCGGACCGCATGGCTACCGAG 3359
DB  ArgArgGluValThrTyrLeuGlyTyrSerLeuArgAspGlyGlnArgTrpLeuThrGlu 400

QY  3360 GCACGGAAGAAAACTGTAGTCCAGATACCCGGCCCAACACAGCAACAAATCAGAGAG 3419
DB  AlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLysGlnMetArgGlu 420

QY  3420 TTTTGGGGACAGCTGGATTTTCAGACTGTGATCCCGGGTTTCGACCTTTAGCAGCC 3479
DB  PheLeuGlyThrAlaGlyPheCysArgLeuTyrIleProGlyPheAlaThrLeuAlaAla 440

QY  3480 CCATCTACCCGCTAACCAAGAAAAAGGGAAATTTCTCTGGGCTCTCTGAGCACCAGAAG 3539
DB  ProLeuTyrProLeuThrLysGluLysGlyGluPheSerTrpAlaProGluHisGlnLys 460

QY  3540 GCATTTGATGCTATCAAAAGAGCCCTGTGAGCGCACCTCTCTGGGCCCTCCCTGACGTA 3599
DB  AlaPheAspAlaIleLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspVal 480

QY  3600 ACTAAACCCCTTTACCTTTATGTGATGACGCGTAAGGGAGTAGCCGGGGAGTTTAAACC 3659
DB  ThrLysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValLeuThr 500

QY  3660 CAAACCTAGGACCATGGAGAACCTGTGCTCTACTCTGTCAAAGAAGCTCGATCTGTGA 3719
DB  GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLeuAspProVal 520

QY  3720 GCCAGTGGTTGGCCCATATGCTGAAGGCTATCGCAGCTGTGCGCATCTACTGTCAAAGAAGCTCGATCTGTGA 3779
DB  AlaSerGlyTrpProLysLysAlaIleAlaValAlaIleLeuValLysAsp 540

QY  3780 GCTGACAAATTGACTTTGGGACAGAATATAACTGTAACTAGTATAGCCCCCATGATTGGAGAAC 3839
DB  AlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProHisAlaLeuGluAsn 560

QY  3840 ATCGTTGGCAGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCCACTATCAAGC 3899
DB  IleValArgGlnProProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSer 580

QY  3900 CTGCTTCTCACAGAGGGGTACGTTTCGCTCCACACAGCGCTCTCAACCTGCCACTCTT 3959
DB  LeuLeuLeuThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600

QY  3960 CTGCTTGAAGAGACTGATGAACCAAGTCACTCATGATTGGCATCACTATTGATTGAGGAG 4019
DB  LeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLeuIleGluGlu 620

QY  4020 ACTGGGGTCCGAGGACCTTACAGACATACCGCTGACCTGGAGAAAGTCTTAACCTGGTTC 4079
DB  ThrGlyValArgLysAspLeuThrAspIleProLeuThrGlyGluValLeuThrTrpPhe 640

QY  4080 ACTGACGGAAGCAGCTATGTGTGGAGGTAAAGAGGATCGCTGGCGCGGGTGGAC 4139
DB  ThrAspGlySerSerTyrValValGluGlyLysArgMetAlaGlyAlaValValAsp 660

QY  4140 GGGACCCGACGATCTGGGCCAGCAGCTGCCGGAAGGAATCTTCAGCACAAAAGGCTGAG 4199
```

Db 661 GlyThrArgThrIleTTPAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 680  
Qy 4200 CTATGCGCCCTCAGCAGAGCTTTGGCTGGCCCAAGGAAATCCATAAATTTATACG 4259  
Db 681 LeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAasnIleTyrThr 700  
Qy 4260 GACAGCAGGTATGCTTTGGACATGCACACTACATGGGGCATCTATAACAAGGGGG 4319  
Db 701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrLysGlnArgGly 720  
Qy 4320 TTGCTTACCTCAGCAGGAGGAAATAAGAAACAAAGAGAAATCTTAAGCCTATTAGAA 4379  
Db 721 LeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluIleLeuSerLeuLeuGlu 740  
Qy 4380 GCGCTACATTACCAAAAGGCTAGCTATTATACACTGTCTCTGCACATCAGAAAGCTAAA 4439  
Db 741 AlaValHisLeuProLysArgLeuAlaIleIleHisCysProGlyHisGlnLysAlaLys 760  
Qy 4440 GATCTCATATCCAGAGNAACCATGGCTGACCGGTTGCCAAGCAGGAGCCAGGCT 4499  
Db 761 AspLeuIleSerArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaGlnGly 780  
Qy 4500 GTTAACCTTCTGCTATAATAGAAATGCCAAAGCCCAAGCCAGACCCAGACGACGTACAC 4559  
Db 781 ValAsnLeuLeuProIleIleGluMetProLysAlaProGluProArgGlnTyrThr 800  
Qy 4560 CTAGAAGACTGGCAAGAGATAAAGAGATAGACAGACTTCTGTAGACTCCGGAGGAGACC 4619  
Db 801 LeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGlyThr 820  
Qy 4620 TGCTATACCTCAGATGGGAAGAAATCTGCGCCCAAGAAAGGTTAGATATGTCCAA 4679  
Db 821 CysTyrThrSerAspGlyLysGluIleLeuProHisLysGluGlyLeuGluTyrValGln 840  
Qy 4680 CAGATACATCTCTAAACCCACTAGGAATAAACACCTGCAGCAGTGGTGGTCAGAACATCC 4739  
Db 841 GlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGlnLeuValArgThrSer 860  
Qy 4740 CCTATCATGTTCTGAGCTACAGGAGTGCTGACTCGGTGGTCAACATTTGTGGCC 4799  
Db 861 ProTyrHisValLeuArgLeuProGlyValAlaAspSerValLysHisCysValPro 880  
Qy 4800 TGCCAGCTGGTTAATGCTAATCTCTCCAGATGCTCCAGGGAAGAGACTAAGGGAGC 4859  
Db 881 CysGlnLeuValAsnAlaAsnProSerArgMetProProGlyLysArgLeuArgGlySer 900  
Qy 4860 CACCCAGCGCTCACTGGGAAGTGGACTTCACTCAGAGTAAAGCCGGCTAAATACGGAAC 4919  
Db 901 HisProGlyAlaHisTTPGluValAspPheThrGluValLysProAlaLysTyrGlyAsn 920  
Qy 4920 AAATACCTATTGTTTGTAGACACCTTTTCAGATGGGTAGAGGCTTATCTCTACTAAG 4979  
Db 921 LysTyrLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaTyrProThrLys 940  
Qy 4980 AAAGAGACTTCAACCGTGGCTTAAATAAATACTGGNAGAAATTTTCCAGATTTCGA 5039  
Db 941 LysGluThrSerThrValValAlaLysLysIleLeuGluGluIlePheProArgPheGly 960  
Qy 5040 ATACTAAGGTAATAGGTCAGACAATGGTCCAGCTTTTGTGTGCCAGGTAAAGTCAGGGA 5099  
Db 961 IleProLysValIleGlySerAspAsnGlyProAlaPheValAlaGlnValSerGlnGly 980  
Qy 5100 CTGGCCAAAGATATTGGGGATTGATTGGAACTGCATTTGTGCATACAGACCCCAAGCTCA 5159  
Db 981 LeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSer 1000  
Qy 5160 GGACAGGTAGAGAGATGAATAGAACCATTAAGAGACCTTACTTAATTCACCGCGAG 5219  
Db 1001 GlyGlnValGluArgMeCAsnArgThrIleLysGluThrLeuThrLysLeuThrAlaGlu 1020  
Qy 5220 ACTGCGCTTAATGATTGATAGTCTCTCGCCCTTTGTGCTTTTGTAGGGTTAGGAACACC 5279

Db 1021 ThrGlyValAsnAspTrpIleAlaLeuLeuProPheValLeuPheArgValArgAsnThr 1040  
Qy 5280 CCTGGACAGTTTGGCTGACCCCTATGAATTACTCTACCGGGACCCCCCAATTTGTA 5339  
Db 1041 ProGlyGlnPheGlyLeuThrProTyrGluLeuLeuTyrGlyGlyProProLeuVal 1060  
Qy 5340 GAAATCTCTCTGTACATAGTGTGCTGCTTTCCAGCCCTTTGTTCTCTAGGCTC 5399  
Db 1061 GluIleAlaSerValHisSerAlaAspValLeuLeuSerGlnProLeuPheSerArgLeu 1080  
Qy 5400 AAGCACTTGAAGTGGTGAGAACAGCGGTGGAGCAACTCCGGGAGGCTACTCAGGA 5459  
Db 1081 LysAlaLeuGluTTPValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSer--- 1099  
Qy 5460 GGAGGACACTTGCAGATCCCATCTGTTCCNAGTGGGAGATTTCAGTCTAGTTAGGCG 5519  
Db 1100 GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTyrValArg 1119  
Qy 5520 CACCGTCAGGAAACCTCGAGACTCGGTGGAGGCGCTTATCTCTGACTTTTTCACACA 5579  
Db 1120 HisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProTyrLeuValLeuLeuThrThr 1139  
Qy 5580 CCAACGCTGTGAAGTCGAAGGAATCTCCACCTGGATCCATGATCCAGTTAAACCG 5639  
Db 1140 ProThrAlaValLysValGluGlyIleProThrTrpIleHisAlaSerHisValLysPro 1159  
Qy 5640 GCGCCACCTCCGATTCGGGGTGGAAAGCCGAAAGACTGAAATCCCTTAAAGCTTCG 5699  
Db 1160 AlaProProAspSerGlyTrpLysAlaGluLysThrGluAsnProLeuLysLeuArg 1179  
Qy 5700 CTCCATCGCGTGGTTCCTTACTCTGTCAATAACCTCTCA 5738  
Db 1180 LeuHisArgValValProTyrSerValAsnAsnSerSer 1192

RESULT 7  
Q90RL9\_9GAMR  
ID Q90RL9\_9GAMR PRELIMINARY; PRT; 1193 AA.  
AC Q90RL9;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polymerase (Fragment).  
GN Name=pol;  
OS Porcine endogenous type C retrovirus.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gamma retrovirus.  
OX NCBI\_TaxID=70540;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21256017; PubMed=11356953;  
RX DOI=10.1128/JVI.75.12.5465-5472.2001;  
RA Krach U., Fischer N., Czaderna F., Toenjes R.R.;  
RT "Comparison of replication-competent molecular clones of porcine  
endogenous retrovirus class A and class B derived from pig and human  
cells.";  
RL J. Virol. 75:5465-5472(2001).  
CC -; SIMILARITY: Contains 1 reverse transcriptase domain.  
DR EMBL; AJ293656; CAC39617.1; -; Genomic\_DNA.  
DR HSSP; P03355; 116J.  
DR MEROPS; A02.020; -.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; F:RNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR009007; Pept\_Aspartc\_cat.  
DR InterPro; IPR001995; Peptidase\_A2\_cat.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR012337; RNaseH\_fold.

DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT; 1.  
DR PROSITE; PS50175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS50994; INTEGRASE; 1.  
DR PROSITE; PS50879; RNASE H; 1.  
DR PROSITE; PS50878; RT POL; 1.  
KW Aspartyl protease; Protease.  
FT NON TRS  
SQ SEQUENCE 1193 AA; 133225 MW; 8FC1F419CB600332 CRC64;  
  
Alignment Scores:  
Pred. No.: 0 Length: 1193  
Score: 6170.50 Matches: 1168  
Percent Similarity: 98.9% Conservative: 12  
Best Local Similarity: 97.9% Mismatches: 12  
Query Match: 42.2% Indels: 1  
DB: 2 Gaps: 1  
  
US-10-723-552-3 (1-8132) x Q90RL9\_9GMR (1-1193)  
  
QY 2160 GGGAGAGCGGGTTCGGACCCCTCCCGAGCCCGAGGTAACTTTGAAGGTGGAGGGCAA 2219  
DB 1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysValGluGlyGln 20  
  
QY 2220 CCAGTTCAGTTCCTGGTTGATACCGGAGCGGAAACATTTCAGTGTCTACTACAGCCATTAGGA 2279  
DB 21 ProValGluPheLeuValAspThrGlyAlaLysHisSerValLeuLeuGlnProLeuGly 40  
  
QY 2280 AAACTAAAGATAAAAAATCTCGGTGATGGGTGCCACAGGGCAAACAACAGTATCCATGG 2339  
DB 41 LysLeuLysAspLysLysSerTipValMetGlyAlaThrGlyGlnGlnGlnTyrProTyr 60  
  
QY 2340 ACTACCCGAGAACAGTGTGCTGGAGTGGAGCGGTAAACCCACTCGTTCTGTGTCATA 2399  
DB 61 ThrThrArgArgThrValAspLeuGlyValGlyArgValThrHisSerPheLeuValIle 80  
  
QY 2400 CCTGAGTGCACGACCCCTTAGGTAGAGACTTATTGACCAAGATGGGAGCACAATTT 2459  
DB 81 ProGluCysProAlaProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100  
  
QY 2460 TCTTTTCAACAAAGGGAACAGAGTGTCTGCAATAACAAACCTATCACTGTGTGTCACC 2519  
DB 101 SerPheGluGlnGlyLysProGluValSerAlaAsnAsnLysProIleThrValLeuThr 120  
  
QY 2520 CTCGAATTAGATGACGAATATCGATATACCTCTCCCTAGTAAAGCCGTGATCAAAATATA 2579  
DB 121 LeuGlnLeuAspAspGluTyrArgLeuTyrSerProLeuValLysProAspGlnAsnIle 140  
  
QY 2580 CAATTCGTGTGGNACAGTTTCCCAAGCTGGCGAGAAACCGCAGGGATGGGTTGGCA 2639  
DB 141 GlnPheTyrLeuLeuGlnPheProGlnAlaTyrPalaGluThrAlaGlyMetGlyLeuAla 160  
  
QY 2640 AAGCAAGTCTCCCCACACAGTTATTCAACTGAAGCCAGTGCCACACAGGTGTCAAGTCAGA 2699  
DB 161 LysGlnValProProGlnValIleGlnLeuLysAlaSerAlaProValSerValArg 180  
  
QY 2700 CAGTACCCCTTGATGAAGAGCTCAAGAAGGAATTCGGCCGCGATGTCCAAAGATTAAATC 2759  
DB 181 GlnTyrProLeuSerLysGluAlaArgGluGlyIleArgProHisValGlnArgLeuIle 200  
  
QY 2760 CAACAGGGATCCTAGTTCCTGTGCAATCTCCCTGGAACTACTCCCTGTCTACCGGTTAGA 2819  
DB 201 GlnGlnGlyIleLeuValProValGlnSerProTyrAsnThrProLeuLeuProValArg 220  
  
QY 2820 AAGCCTGGGACTAATGACTATCGACCAAGTACAGACTTGAGAGAGGTCAATAAACGGGTG 2879  
DB 221 LysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsnLysArgVal 240  
  
QY 2880 CAGGATATACACCAACAGTCCCGAACCCCTTATAACCTCTTGTGTCTCTCCCAACCCCAA 2939

DB 241 GlnAspIleHisProThrValProAsnProTyrAsnLeuLeuCysAlaLeuProGln 260  
QY CCGAGCTGTATACAGTATTGGACTTAAAGATCCCTTCTCTCCCTGAGATTACACCCC 2999  
DB 261 ArgSerTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisPro 280  
QY 3000 ACTAGCCAAACACCTTTTTCCTTCGAATGGAGAGATCCAGTACGGGAAGAACCGGGCAG 3059  
DB 281 ThrSerGlnProLeuPheAlaPheGluTyrArgAspProGlyAlaGlyArgThrGlyGln 300  
QY 3060 CTCACCTGGACCCGACTGCCCCAAAGGTTTCAAGAACTCCCGACCATCTTTTACGAAAGCC 3119  
DB 301 LeuThrTyrThrArgLeuProGlnGlyPheLysAsnSerProThrIlePheAspGluAla 320  
QY 3120 CTACACAGAGACCTGGCCAACTTCAGGATCCACACCTCAGGTGACCTCTCCCTCCAGTAC 3179  
DB 321 LeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnTyr 340  
QY 3180 GTGGATGACCTGTCTTTCGGCGGAGCCACCAACAGGACTGTCTAGAAGGACCAAGGCA 3239  
DB 341 ValAspAspLeuLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGluGlyThrLysAla 360  
QY 3240 CTACTGCTGAATTGTCTGACCTAGGCTACAGAGCTCTCTGCTAAGAAGGCCAGATTGTC 3299  
DB 361 LeuLeuLeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLysAlaGlnIleCys 380  
QY 3300 AGGAGAGAGTAACATCTTGGGTGACAGTTTGGCGGACGGGACGCGATGGCTGACGGAG 3359  
DB 381 ArgArgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlnArgTyrLeuThrGlu 400  
QY 3360 GCACGGGAAGAAACTGTAGTCCAGATACCGGCCCAACACAGCAACCAAAATCAGAGAG 3419  
DB 401 AlaArgLysArgThrValValGlnIleProAlaProThrThrAlaLysGlnValArgGlu 420  
QY 3420 TTTTGGGACAGCTGGATTTGCAGACTGTGATCCCGGGTTTGGACCTTTAGCAGCC 3479  
DB 421 PheLeuGlyThrAlaGlyPheCysArgLeuTyrPheLeuAlaThrLeuAlaAla 440  
QY 3480 CCATCTACCCGCTAACCAAGAAAAGGGAATTTCTCTGGCTCTCTGAGCACCAGAAC 3539  
DB 441 ProLeuTyrProLeuThrLysGluLysGlyGluPheSerTyrPalaProGluHisGlnLys 460  
QY 3540 GCATTTGATCTATCAAAAGGCCCTGTCTGAGCGCACCTGTCTGGGCCCTCCCTGAGCTA 3599  
DB 461 AlaPheAspAlaIleLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspVal 480  
QY 3600 ACTAAACCCCTTTTACCTTTATGTGATGAGCGTAAAGGAGTAGCCCGGGAGTTTAAACC 3659  
DB 481 ThrLysProPheThrLeuTyrValAspGluArgLysGlyValAlaArgGlyValLeuThr 500  
QY 3660 CAACCCCTAGGACCATGGAGAACCTGTGCGCTACTCTGTCAAGAAGAGCTCGATCTCTGTA 3719  
DB 501 GlnThrLeuGlyProTyrArgArgProValAlaTyrLeuSerLysLysLeuAspProVal 520  
QY 3720 GCCAGTGTGTGGCCCATATGCTGAAGGCTATCGCAGCTGTGGCTGCTACTGTGTCAGGAC 3779  
DB 521 AlaSerGlyTyrProLysLysLysAlaIleAlaValAlaIleLeuValLysAsp 540  
QY 3780 GCTGACAAATTTGACTTTTGGGACAGAATATAACTGTATATAGCCCCCATGATCTTGGAGAAC 3839  
DB 541 AlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProHisAlaLeuGluAsn 560  
QY 3840 ATGCTTGGCAGCCCCCAGACCGATGATGACCAACGCGCCGATGACCCACTATCAAGC 3899  
DB 561 IleValArgGlnProProAspArgTyrMetThrAsnAlaArgMetThrHisTyrGlnSer 580  
QY 3900 CTGCTTCTCAGCAGAGAGGTTCAGTTTCGCTCCACAGCGCTCTCAACCTGCGCACTCTT 3959  
DB 581 LeuLeuLeuThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600  
QY 3960 CTGCTTGAAGAGACTGATGAACCAAGTCTCATGATGTCATGTCATCACTATTGATTGGAG 4019



Db 601 LeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLeuLeuGluGlu 620  
Qy 4020 ACTGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGAGAGAGTGCCTAACCTGGTTC 4079  
Db 621 ThrGlyValArgLysAspLeuThrAspIleProLeuThrGlyGluMetLeuThrTrpPhe 640  
Qy 4080 ACTCAGGGAAGCAGCATGTGCTGGAAGTAAAGAGGATGGCTGGGGCGGCTGGTGGAC 4139  
Db 641 ThrAspGlySerSerTrpMetValGluGlyLysArgMetAlaGlyAlaIvalValAsp 660  
Qy 4140 GGGACCCGACGATCTGGGCCAGCAGCTGCCGAAGAACTTCAGCACAAAAGCTGAG 4199  
Db 661 GlyThrArgThrIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 680  
Qy 4200 CTCATGGCCCTCAGCAAGCTTTGGGCTGGCCGAGGGAATCCATAAACAATTTATACG 4259  
Db 681 LeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTrpThr 700  
Qy 4260 GACAGCAGGTATGCTTTGGCACTGCACACGTACATGGGGCCATCTATAACAAGGGGG 4319  
Db 701 AspSerArgTrpAlaPheAlaThrAlaHisValHisGlyAlaIleTrpLysGlnArgGly 720  
Qy 4320 TTGCTTACTCAGCAGGGAGGAAATAAGAACAAAGAGGAAATCTTAAGCCTATTAGAA 4379  
Db 721 LeuLeuThrSerAlaGlyArgGluIleLysAsnLysGlnGluIleLeuSerLeuLeuGlu 740  
Qy 4380 GCCGTACATTTACAAAAGGCTAGCTATTATATACATGCTCTGACATCATCAAGAGCTAAA 4439  
Db 741 AlaLeuHisLeuProLysArgLeuAlaIleHisCysProGlyHisGlnLysAlaLys 760  
Qy 4440 GATCTCATATCCAGAGGAACACAGATGCTCACCGGTTGCAAGCAGCGAGCCAGGGT 4499  
Db 761 AspProIleSerArgLysGlnMetAlaAspArgValAlaLysGlnAlaIleGlnGly 780  
Qy 4500 GTTAACCTTCTGCTTATAATAGAATGCCAAAGCCCAAGCCCAAGCAGCAGCAGTACAC 4559  
Db 781 ValAsnLeuLeuProMetIleGluThrProLysAlaProGluProGlyArgGlnTrpThr 800  
Qy 4560 CTAGAAGACTGGCAAGAGATAAAAGATAGACACAGTTCTCTGAGACTCCGGAAGGGACC 4619  
Db 801 LeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGlyThr 820  
Qy 4620 TGCTATACCTCAGATGGGAAGAAATCCTGCCCAAGAAAGGTTAGATATGTCCAA 4679  
Db 821 CysTrpThrSerAspGlyLysGluIleLeuProHisLysGluGlyLeuGluTrpValGln 840  
Qy 4680 CAGATACATCTCTAACCCACCTAGGAACCTGACAGCAGTGGTGGTACAGAACATCC 4739  
Db 841 GlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGlnLeuValArgThrSer 860  
Qy 4740 CTTATCATGTTCTGAGGCTACCAGGAGTGGCTGACTGGTGGTCAACATTTGTGCCC 4799  
Db 861 ProTrpHisValLeuArgLeuProGlyValAlaAspSerValValLysHisCysValPro 880  
Qy 4800 TGCCAGCTGGTTAATGCTAATCTCTCCAGAATGCTCCAGGGAAGAGACTAAGGGGAAGC 4859  
Db 881 CysGlnLeuValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySer 900  
Qy 4860 CACCAGCGCTCACTGGGAAGTGGACTTCCTCAGAGTAAAGCCGGCTAAATACGGAAC 4919  
Db 901 HisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysTrpGlyAsn 920  
Qy 4920 AAAATACCTATTGGTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCTCTAAG 4979  
Db 921 LysTrpLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaLysProThrLys 940  
Qy 4980 AAAGAGACTCAACCGTGGTCTAAAATAATCTCGAAGAAATTTTCCAGATTTGGA 5039  
Db 941 LysGluThrSerThrValValAlaLysLysIleLeuGluGluIlePheProArgPheGly 960  
Qy 5040 ATACCTAAGGTAATAGGCTCAGACAAATGGTCCAGCTTTTGTGGCCAGGTAAAGTCAGGA 5099  
Db 961 IleProLysValIleGlySerAspAsnGlyProAlaPheValAlaGlnValSerGlnGly 980

Qy 5100 CTGCCAAGATATTGGGATTGATTGGAACTGTCATTGTGCATACACACCCCAAGCTCA 5159  
Db 981 LeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTrpArgProGlnSerSer 1000  
Qy 5160 GGCAGCTAGAGAGGATGAATAGAACCATTAAGAGACCCCTTACTATAATTGACCGGAG 5219  
Db 1001 GlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThrGlu 1020  
Qy 5220 ACTGGCGTTAATGATTGGATAGCTCTCTCCCTTTGTGCTTTTGTAGGGTTAGAAACAC 5279  
Db 1021 ThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeuPheArgValArgAsnThr 1040  
Qy 5280 CCTGCAGAGTTTGGCTGACCCCTATGAATTACTACGGGGAGCCCCCCCAATTGGTA 5339  
Db 1041 ProGlyGlnPheGlyLeuThrProTrpGluLeuLeuTrpGlyGlyProProProLeuAla 1060  
Qy 5340 GAAATGCTTCTGTACATAGTGTGACGTCTGCTTCCAGCGCTTTTGTCTTAGGCTC 5399  
Db 1061 GluIleAlaPheAlaHisSerAlaAspValLeuLeuSerGlnProLeuPheSerArgLeu 1080  
Qy 5400 AAGCATTCTGAGTGGTGACAAACAGCGTGGAGGCAACTCCGGGAGGCTACTCAGGA 5459  
Db 1081 LysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTrpSer 1099  
Qy 5460 GGAGGACACTTGCAGATCCCAATCGTTTCCAGTGGGAGATTTCAGTCTACGTTAGACGC 5519  
Db 1100 GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTrpValArg 1119  
Qy 5520 CACCGTCAGAGAAACCTCGAGACTCGTGGAGGCGCCCTTATCTCGTACTTTTACCACA 5579  
Db 1120 HisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProTrpLeuValLeuLeuThr 1139  
Qy 5580 CCAACGGCTGTGAAGTCGAAGGATCTCCACCTGGATCCATGTCATCCACGTTAAACCG 5639  
Db 1140 ProThrAlaValLysValGluGlyIleProThrTrpIleHisAlaSerHisValLysPro 1159  
Qy 5640 GCGCCACTCCGATTCGGGTGGAAAGCCGAAAGACTGAAATCCCTTAAAGCTTCGC 5699  
Db 1160 AlaProProProAspSerGlyTrpLysAlaGluLysThrGluAsnProLeuLysLeuArg 1179  
Qy 5700 CTCCATCGCTGGTTCCTTACTCTGTCAATAACCTCTCA 5738  
Db 1180 LeuHisArgValValProTrpSerValAsnAsnSerSer 1192  
RESULT 8  
Q90RL6\_9GAMR PRELIMINARY; PRT; 1195 AA.  
AC Q90RL6;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polymerase (Fragment).  
GN Name=pol;  
OS Porcine endogenous type C retrovirus.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gammaretrovirus.  
OX NCBI\_taxid=70540;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21256017; PubMed=11356953;  
RX DOI=10.1128/JVI.75.12.5465-5472.2001;  
RA Krach U., Fischer N., Czaderna F., Toenjes R.R.;  
RT "Comparison of replication-competent molecular clones of porcine  
RT endogenous retrovirus class A and class B derived from pig and human  
RT cells.";  
RL J. Virol. 75:5465-5472(2001).  
CC -I- SIMILARITY: Contains 1 reverse transcriptase domain.  
DR EMBL; AJ293657; CAC39626.1; -; Genomic\_DNA.  
DR HSP; P03355; 116J.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.



DR GO: 0008233; F: peptidase activity; IEA.  
 DR GO: 0004523; F: ribonuclease H activity; IEA.  
 DR GO: 0003723; F: RNA binding; IEA.  
 DR GO: 0003964; F: RNA-directed DNA polymerase activity; IEA.  
 DR GO: 0006310; P: DNA recombination; IEA.  
 DR GO: 0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: 0006278; P: RNA-dependent DNA replication; IEA.  
 DR InterPro: IPR001969; Pept\_Asp\_AS.  
 DR InterPro: IPR009007; Pept\_Aspartc\_cat.  
 DR InterPro: IPR001995; Peptidase\_A2\_cat.  
 DR InterPro: IPR002156; RNaseH.  
 DR InterPro: IPR012337; RNaseH\_fold.  
 DR InterPro: IPR001584; Rve.  
 DR InterPro: IPR000477; RVtse.  
 DR Pfam: PF00075; RNaseH.1.  
 DR Pfam: PF00665; rve.1.  
 DR Pfam: PF00077; RVP.1.  
 DR Pfam: PF00078; RVT.1.  
 DR PROSITE: PS0175; ASP\_PROT\_RETROV.1.  
 DR PROSITE: PS0141; ASP\_PROTEASE.1.  
 DR PROSITE: PS0994; INTEGRASE.1.  
 DR PROSITE: PS0879; RNASE H.1.  
 DR PROSITE: PS0878; RT POL.1.  
 DR Aspartyl protease; Hydrolase; Protease.

FT NON\_TER 1

SQ SEQUENCE 1195 AA; 133770 MW; 999F1524DE0BFFA4 CRC64;

#### Alignment Scores:

Pred. No.:	0	Length:	1195
Score:	6109.00	Matches:	1155
Percent Similarity:	98.6%	Conservative:	19
Best Local Similarity:	97.0%	Mismatches:	17
Query Match:	41.7%	Indels:	0
DB:	2	Gaps:	0

US-10-723-552-3 (1-8132) x Q90RL6\_9GMR (1-1195)

QY	2160	GGGAGACGGGTTCCGACCCCTCCCGAGCCAGGGTAACTTTGAAGTGGAGGGCAA	2219
DB	1	GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysValGluGln	20
QY	2220	CCAGTTGAGTCTGGTTGATACCGGAGCGAAACATTCAGTGTCTACTACAGCCATTAGGA	2279
DB	21	ProValGluPheLeuValAspThrGlyAlaGluHisSerValLeuLeuGlnProLeuGly	40
QY	2280	AAACTAAAGATAAAATCTGGTGCATGGTCCACAGGGCAACACAGTATCCATGG	2339
DB	41	LysLeuLysGluLysLysSerTrpValMetGlyAlaThrGlyGlnArgGlnTrpProTrp	60
QY	2340	ACTACCCGAAGACAGTTGACTTGGGAGTGGGAGTAAACCCACTCGTTTCTGGTCATA	2399
DB	61	ThrThrArgArgThrValAspLeuAlaValGlyArgValThrHisSerPheLeuValIle	80
QY	2400	CCTGAGTGGCCAGCACCCCTCTAGGTAGAGACTTATGACCAAGATGGGAGCACAAATT	2459
DB	81	ProGluCysProValProLeuLeuGlyArgAspLeuThrLysMetGlyAlaGlnIle	100
QY	2460	TCITTTTGAACAGGGAACAGAAAGTCTCGCAATAACAACTACACTGTGTGTGACC	2519
DB	101	SerPheGluGlnArgProGluValSerValAsnAsnLysProIleThrValLeuThr	120
QY	2520	CTCCAAATTAGATGACGAATATCGACTATCTCTCCCTAGTAAAGCCCTGATCAAAATATA	2579
DB	121	LeuGlnLeuAspAspGluTyArgLeuTySerProGlnValLysProAspGlnAspIle	140
QY	2580	CAATTCGTGGTGAACAGTTTCCCAAGCTGGCGAGAAACCGGAGGATGGGTTGGCA	2639
DB	141	GlnSerTrpLeuGluGlnPheProGlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAla	160
QY	2640	AACCAAGTCCCCACACAGTTATTCACTCAAGCCAGTGCACACCAAGTGTGAGTCAGA	2699
DB	161	LysGlnValProProGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArg	180

QY	2700	CAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTCGGCGCATGTCCAAAGATTATTC	2759
DB	181	GlnTyProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgLeuIle	200
QY	2760	CAACAGGGCATCCTAGTTCCTGCTCCATCTCCCTGGAAATCTCCCTGCTACCCGTTAGA	2819
DB	201	GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLeuProValArg	220
QY	2820	AAGCCTCGGACTAATGACTATCGACCACTAGTACAGGACTTGAGAGAGTCAATAAAGCGGTG	2879
DB	221	LysProGlyThrAsnAspTyArgProValGlnAspLeuArgGluValAsnLysArgVal	240
QY	2880	CAGGATATACACCCCAACAGTCCCGAACCTTATAACCTTCTGTGTGCTCTCCACCCCAA	2939
DB	241	GlnAspIleHisProThrValProAsnProTyAsnLeuLeuSerAlaLeuProGlu	260
QY	2940	CGGAGCTGGTATACAGTATTGGACTTAAAGATGCCTTCTTCGCTGAGATTACACCCC	2999
DB	261	ArgAsnTrpTyThrValLeuAspLeuLysAspAlaPheCysLeuArgLeuHisPro	280
QY	3000	ACTAGCCCAACACTTTTTCCTTCGAATCGAGAGATCCAGGTACGGGAAGAACCGGCAG	3059
DB	281	ThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln	300
QY	3060	CTCACCTGGACCCGACTGCCCCAAGGGTTCAGAAGCTCCCGACCATCTTTTGAGAAAGCC	3119
DB	301	LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrPheAspGluAla	320
QY	3120	CTACACAGACACTGGCCAACTTCAGATCCACACCTCAGGTGACCTCTCCAGTAC	3179
DB	321	LeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnTy	340
QY	3180	GTGGATGACCTGCTCTGGCGGAGCCACCAACAGGACTCTTAGAAGCCACGAGCGCA	3239
DB	341	ValAspAspLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGluGlyThrLysAla	360
QY	3240	CTACTCTGGAATGTGTGACTAGGCTACAGAGCTCTCTGAAGAGGCCAGATTTC	3299
DB	361	LeuLeuLeuGluLeuSerAspLeuGlyTyArgAlaSerAlaLysLysAlaGlnIleCys	380
QY	3300	AGGAGAGAGTAACTACTTGGGTACAGTTTCGGGAGCGGAGCGGATGGTACCGAG	3359
DB	381	ArgArgGluValThrTyLeuGlyTySerLeuArgGlyGlnArgTrpLeuThrGlu	400
QY	3360	GCACGGAAGAAAACCTGTAGTCAGATACCGGCCCAACACACAGCAAAACAAATAGAGAG	3419
DB	401	AlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLysGlnValArgGlu	420
QY	3420	TTTTTGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTGCACCTTTAGCAGCC	3479
DB	421	PheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAlaAla	440
QY	3480	CAACTCTACCGGTAAACCAAGAAAAGGGAATTCCTCTGGGCTCTCTAGCAGCAGAG	3539
DB	441	ProLeuTyProLeuThrLysGluLysGlyGluPheSerTrpAlaProGluHisGlnLys	460
QY	3540	GCATTTGATGCTATCAAAAGGCCCTGTGAGCGCACCTCTCTGGCCCTCCCTGAGCTA	3599
DB	461	AlaPheAspAlaLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspVal	480
QY	3600	ACTAAACCCCTTTTACCCTTTTATGTGGATGAGCGTAAAGGAGTAGCCCGGGGAGTTTAAAC	3659
DB	481	ThrLysProPheThrLeuTyValAspGluArgLysGlyValAlaArgGlyValLeuThr	500
QY	3660	CAAAACCTTAGGACCATGGAGAGACCTGTCCGCTACTCTGTCAAAGAGCTCGATCTGTA	3719
DB	501	GlnThrLeuGlyProTrpArgArgProValAlaTyLeuSerLysLysLeuAspProVal	520
QY	3720	GCCAGTGGTGGCCCATATGCTGAAGGCTATCGAGCTGTGGCCATCTGTCAGTCGTA	3779
DB	521	AlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaValAlaIleLeuValLysAsp	540
QY	3780	GCTGACAAATTGACTTTTGGGACAGAAATATAACTGTAAATAGCCCCCATTCATTTGGAGAAC	3839

Db 541 AlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProHisAlaLeuGluAsn 560  
Qy 3840 ATCGTTGGCAGCCCCAGACCGATGGATGACCAACGCCCGCATAGCCACTATCAAGC 3899  
Db 561 IleValArgGlnProProAspArgTrpMetThrAsnAlaArgMetThrHisTyGlnSer 580  
Qy 3900 CTGCTTCTCAGAGAGGGTCACGTTGGCTCCACAGCGCTCTCAACCTGGCCACTCTT 3959  
Db 581 LeuLeuLeuThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600  
Qy 3960 CTGCTCAGAGACTGATGAACCACTGACTCATGATTGCCATCAACTATTGATTGAGGAG 4019  
Db 601 LeuProGluGluThrAspGluProValThrHisaspCysHisGlnLeuLeuIle6luGlu 620  
Qy 4020 ACTGGGTCGCCAAGGACCTTACAGACATACCGCTGACTGGAGAGTGCTAACTGGTTC 4079  
Db 621 ThrGlyValArgLysAspLeuThrAspIleProLeuThrGlyGluValLeuThrTrpPhe 640  
Qy 4080 ACTACCGAGACCTATGTGTGGTGAAGGTAAAGGATGGCTGGGGCGGGTGGTGAC 4139  
Db 641 ThrAspGlySerSerTyTyValValGluGlyLysLysMetAlaGlyAlaAlaValValAsp 660  
Qy 4140 GGGACCCGACGATCTGGGCCAGCAGCTGCGCGAAGAACTTCAGCACAAAAGCGTGAG 4199  
Db 661 GlyThrArgThrIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 680  
Qy 4200 CTGATGGCCCTCAGCAGCTTTGGCTGGCCGCAAGGAAATCCATAAATTTATACG 4259  
Db 681 LeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyThr 700  
Qy 4260 GACAGCAGGTATGCTTTGGACTGCACGTACATGGGCGCATCTATAACAAAGGGG 4319  
Db 701 AspSerArgTyAlaPheAlaThrAlaHisValHisGlyAlaIleTyLysGlnArgGly 720  
Qy 4320 TTGCTTACTCAGCAGGAGGGAATAAAGAACAAAGAGAAATCTTAAGCCTATTAGAA 4379  
Db 721 LeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluIleLeuSerLeuLeuGlu 740  
Qy 4380 GCCGTACATTTACAAAAGGCTAGCTATTATACATCTCTGTGACATCAAGAAAGCTAAA 4439  
Db 741 AlaLeuHisLeuProLysArgLeuAlaIleIleHisCysProGlyHisGlnLysAlaLys 760  
Qy 4440 GATCTCATATCCAGGNAACAGATGGCTGCTACCGGTTGCGACGAGCGGCCAGGGT 4499  
Db 761 AspLeuIleSerArgLysAsnGlnMetAlaAspArgValAlaLysGlnAlaAlaGlnAla 780  
Qy 4500 GTTAACCTTCTGCTATAATAAGAAATGCCAAAGCCCCAGAACCCAGACGACAGTACACC 4559  
Db 781 ValAsnLeuLeuProIleIleGluThrProLysAlaProGluProArgArgGlnTyThr 800  
Qy 4560 CTAGAAGACTGGCAAGAGATAAAAGATAGACCACTGCTCTGAGACTCGGAAGGACCC 4619  
Db 801 LeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGlyThr 820  
Qy 4620 TGCTATACCTCAGATGGGAAGAAATCTGCGCCACAAAGNAGGGTTAGATATGTCAA 4679  
Db 821 CysTyThrSerTyArgLysGluIleLeuProHisLysGluGlyLeuGluTyValGln 840  
Qy 4680 CAGATACATCTCTAAACCCACTAGGAATAAACACCTGCAGCAGTGGTGTCAGAACATCC 4739  
Db 841 GlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGlnLeuValArgThrSer 860  
Qy 4740 CCTTATCATGTTCTGAGGCTACAGGAGTGCTGACTCGGTGGTCAACATTTGTGCCC 4799  
Db 861 ProTyHisValLeuArgLeuProGlyValAlaAspSerValValLysHisCysValPro 880  
Qy 4800 TGCCAGCTGGTTAATGCTAATCTCCAGATGCTCCAGGAGAGACTAAGGGGAGC 4859  
Db 881 CysGlnLeuValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySer 900  
Qy 4860 CACCAGCGCTCACTGGGAAGTGAGCTTCACTCAGGTAAAGCGGCTAAATACGGAAC 4919

Db 901 HisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysTyThrGlyAsn 920  
Qy 4920 AATACCTATTGTTTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACTAAG 4979  
Db 921 LysTyThrLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaTyProThrLys 940  
Qy 4980 AAAGAGACTTCAACCGTGGTGGCTAAAAAAATACTCGAAGAAAATTTTTCAGAAATTTGGA 5039  
Db 941 LysGluThrSerThrValValAlaLysLysIleLeuGluGluIlePheProArgPheGly 960  
Qy 5040 ATACCTAAGCTAATAGGGTCAGACAATGGTCCAGCTTTTGTGCCAGGTAAAGTCAGGA 5099  
Db 961 IleProLysValIleGlySerAspAsnGlyProAlaPheValAlaGlnValSerGlnGly 980  
Qy 5100 CTGGCCAAAGATATTGGGGATTGATGGAAAATGCTGATTTGTCATACAGACCCCAAGCTCA 5159  
Db 981 LeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyArgProGlnSerSer 1000  
Qy 5160 GGACAGCTAGAGAGATGAATAGAACCATTAAGAGACCTTACTAAATTTGACCGCGAG 5219  
Db 1001 GlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThrGlu 1020  
Qy 5220 ACTGGCGCTTAATGATGGATAGCTCTCCTGCCCTTTGTGCTTTTAGGGTTAGGAACACC 5279  
Db 1021 ThrGlyIleAsnAspTrpIleAlaLeuLeuProPheValLeuPheArgValArgAsnThr 1040  
Qy 5280 CTGGACAGTTTGGCTGACCCCTATGAATTAATCTACCGGGGACCCCGCCATTTGGTA 5339  
Db 1041 ProGlyGlnPheGlyLeuThrProTyArgGluLeuLeuTyArgGlyProProLeuVal 1060  
Qy 5340 GAAATGCTTCTGTACATAGTGTGACGTGCTGCTTTCCAGACCTTTGTTCTCTAGGCTC 5399  
Db 1061 GluIleAlaSerValHisSerAlaAspValLeuLeuSerGlnProLeuPheSerArgLeu 1080  
Qy 5400 AAGCAGCTTCAAGTGGGTGACAAACGAGCTGTGAGGCAACTCCGGGAGGCTACTCAGGA 5459  
Db 1081 LysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTySerGly 1100  
Qy 5460 GAGGAGACTTGCAGATCCACATCGTTTCCAAAGTGGGAGATTCAGTCTAGTTAGAGC 5519  
Db 1101 GluGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTyValArgArg 1120  
Qy 5520 CACGCTCAGAAACCTCGAGACTCGGTGGAAGGCCCTTATCTCGTACTCTTTGACCACA 5579  
Db 1121 HisHisLagLysLeuLeuGluThrArgTrpLysGlyProTyLeuValLeuLeuThrThr 1140  
Qy 5580 CCAACGCTGTGAAGTCGAAGGAATCTCCACCTGGATCCATGCATCCACGTTAAACCG 5639  
Db 1141 ProThrAlaValLysValGluGlyIleSerThrTrpIleHisAlaSerHisValLysLeu 1160  
Qy 5640 GCGCCACCTCCGATTCGGGGTGGMAAGCCGAAAAGACTGAAAATCCCTTAAGCTTCGC 5699  
Db 1161 AlaProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeuLysLeuArg 1180  
Qy 5700 CTCCATCGCTGGTTCCTTACTCTGTCAATAAC 5732  
Db 1181 LeuHisArgLeuValProTySerAsnAsnAsn 1191

RESULT 9

Q9XSN8.PIG  
ID O9XSN8.PIG PRELIMINARY; PRT; 1145 AA.  
AC O9XSN8;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Pol protein.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
OC Sus.  
OX NCBI\_TaxID=9823;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.

RC STRAIN=miniature swine;  
RX MEDLINE=98216827; PubMed=9557749;  
RA Akiyoshi D.E., Denaro M., Zhu H., Greenstein J.L., Banerjee P.,  
RA Fishman J.A.;  
RT "Identification of a full-length cDNA for an endogenous retrovirus of  
RT miniature swine."; 4507(1998).  
RL J. Virol. 72:4503-4507(1998).  
DR EMBL; AF038600; AAC16764.1; -; mRNA.  
DR HSSP; P03355; 116J.  
DR MEROPS; A02.020; -.  
DR GO; GO:0004190; P:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; P:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR001995; Peptidase\_A2\_cat.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; Rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR PROSITE; PS0175; ASP PROT RETROV; 1.  
DR PROSITE; PS0879; RNASE H; 1.  
SQ SEQUENCE 1145 AA; 128010 MW; 90C1A95E9D95B88 CRC64;

Alignment Scores:  
Pred. No.: 0 Length: 1145  
Score: 6026.00 Matches: 1143  
Percent Similarity: 99.8% Conservative: 0  
Best Local Similarity: 99.8% Mismatches: 2  
Query Match: 41.2% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3 (1-8132) x Q9XSN8\_PIG (1-1145)

QY 2307 ATGGTGCCACAGGCGACACAGTATCCATGACTACCCGAGACAGTGTGCTGGGA 2366  
Db 1 MetGlyAlaThrGlyGlnGlnTyrProTrpThrThrArgArgThrValAspLeuGly 20

QY 2367 GTGGGCGGTAAACCCACTCGTTCTCGTCTATCTGCTGCTGAGTCCAGCCACCCCTCTTAGGT 2426  
Db 21 ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 40

QY 2427 AGAGACTTATGACCAAGATGGGAGCAAAATTTCTTTGAACAAGGAAACAGAGTG 2486  
Db 41 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyLysProGluVal 60

QY 2487 TCTGCAATTAACAACTATCAGTGTGACCTCCAACTAGATGACGAATATCGACTA 2546  
Db 61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu 80

QY 2547 TACTCTCCCTAGTAAAGCTGTATCAAAATATATCAATTTCTGTTGGAACAGTTTCCCAA 2606  
Db 81 TyrSerProLeuValLysProAspGlnAsnIleGlnPheTrpLeuGlnPheProGln 100

QY 2607 GCTGGGCGAAGAACCGCAGGATGGGTTTGGCAAGCAAGTTCCTCCCAAGTTATTCAA 2666  
Db 101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 120

QY 2667 CTGAAGCCAGTGCACACAGTGTGTCAGACAGTACCCCTTGATGAAGAGCTCAA 2726  
Db 121 LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerLysGluAlaGln 140

QY 2727 GAAGGAATTCGGCCGATGTCCAAAGATTAAATCCACAGGCGCATCTAGTTCCTGTCCTCAA 2786  
Db 141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 160

QY 2787 TCTCCCTGGGAATACTCCCTGCTACCGGTTAGAAAGCCTGGGACTAATGACTATCGACCA 2846

Db 161 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 180  
QY 2847 GTACAGACTTGAGAGAGGTCAATAAACGGGTGAGGATATACACCAACAGTCCGGAAC 2906  
Db 181 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200  
QY 2907 CCTTATAACCTCTGTGTGCTCTCCACCCCAACGGAGCTGTATACAGTATTGAGCTTA 2966  
Db 201 ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTrpTyrThrValLeuAspLeu 220  
QY 2967 AAGATCCCTTCTCTGCTGAGATTACCCCACTAGCCCAACACCTTTTTCCTTCGAA 3026  
Db 221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240  
QY 3027 TGGAGAGATCCAGGTACGGAGAACCGGACGCTCACTCGGACCCGACCTGCCCCAAGG 3086  
Db 241 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260  
QY 3087 TTCAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGG 3146  
Db 261 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280  
QY 3147 ATCCAAACCTCAGGTGACCTCTCCAGTACGTGGATCCTGCTTCTGCGGGAGCC 3206  
Db 281 IleGlnHisProGlnValThrLeuLeuGlnTyr-ValAspAspLeuLeuAlaGlyAla 300  
QY 3207 ACCAAACAGACTGCTTAGAAGGACGACGACTACTGCTGGNATTTGCTGACCTAGGC 3266  
Db 301 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 320  
QY 3267 TACAGAGCTCTGCTAAGAGGCGCCAGATTGACGAGAGAGGTAAACATCTTGGGGTAC 3326  
Db 321 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 340  
QY 3327 AGTTTGGGACGCGGACGCGATGCTGACGAGGACGCGAGGAGAACTAGTTCAGATA 3386  
Db 341 SerLeuArgAspGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 360  
QY 3387 CCGGCCCCAACACACGACCAACAANTAGAGAGATTTTGGGGACAGCTGATTTGGAGA 3446  
Db 361 ProAlaProThrThrAlaLysGlnMetArgGluPheLeuGlyThrAlaGlyPheCysArg 380  
QY 3447 CTGTGATCCCGGGTTTGGACCTTAGCAGGCCCTCTACCCGCTAACCAAGAAAGAAA 3506  
Db 381 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysLys 400  
QY 3507 GGGGAATTCTCTGGGCTCTCTGAGCACCAAGGCAATTTGATGCTATCAAAAAGGCCCTG 3566  
Db 401 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420  
QY 3567 CTGAGGCGACTGCTCTGGCCCTCCCTGAGCTAACTAAACCTTTACCTTTATGTGGAT 3626  
Db 421 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 440  
QY 3627 GAGCGTAAGGAGTACCCGGGAGTTTAAACCAACCCCTAGGACCATCGAGAGACCT 3686  
Db 441 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 460  
QY 3687 GTCGCTACTCTGTCAAAAGAGCTCGATCTCTGTAGCAGGTGTTGGGCCATATGCTCAAG 3746  
Db 461 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLys 480  
QY 3747 GCTATCGAGCTGTGGCCATCTAGTCAAGGACGCTGACAAATGTACTTTGGGACAGAA 3806  
Db 481 AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 500  
QY 3807 ATAACGTATAGCCCCCATGCTATGGAGAACATCGTTCGGCAGCCCCCAGACCCATGG 3866  
Db 501 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 520  
QY 3867 ATGACCAACGCCCGCATGACCCCACTATCAAGCCTCTCTCTCACAGAGGGGTCAAGTTC 3926

Db 521 MetThrAsnAlaArgMetThrHisThrGlnSerLeuLeuLeuLeuThrGluArgValThrPhe 540  
Qy 3927 GCTCACAGCGCTCTCAACCTGCCACTCTTCTGCTGAGAGACTGATGAACCAAGTG 3986  
Db 541 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuLeuProGluGluThrAspGluProVal 560  
Qy 3987 ACTCATGATTCCTCAACTATTGATTGAGGAGACTGGGGTCCCAAGGACTTTACAGAC 4046  
Db 561 ThrHisaspCysHisGlnLeuLeuLeuLeuGluGluThrGlyValArgLysAspLeuThrAsp 580  
Qy 4047 ATACCGCTGACTGAGGAAGTGCTAACCTGGTTCTACTGACGGAAGCAGCTATGTGGTGAA 4106  
Db 581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 600  
Qy 4107 GGTAAAGAGATGGCTGGGGCGGGTGTGACGGGACCCGACGATCTGGCCAGCAGC 4166  
Db 601 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 620  
Qy 4167 CTGCCGGAAGAACTTCAGACACAAAGGCTGAGCTCATGGCCCTCAGCGAGCTTTTCGG 4226  
Db 621 LeuProGlyGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
Qy 4227 CTGCCGGAAGGAAATCCATAAATTTATACGACAGCAGGTATGCCCTTTGCGACTGCA 4286  
Db 641 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 660  
Qy 4287 CACGTACATGGGGCATCTATAACAAAGGGGGTGTCTTACTCTCAGCAGGAGGGAATA 4346  
Db 661 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
Qy 4347 AAGAACAAAGAGGAATTTCTAAGCCTATTACAAGCCGTACATTTTACCAGAAAGGCTAGCT 4406  
Db 681 LysAsnLysGluGluIleLeuSerLeuLeuLeuGluAlaValHisLeuProLysArgLeuAla 700  
Qy 4407 ATTATACACTCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAACACAGATG 4466  
Db 701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720  
Qy 4467 GCTGACCGGGTTGCCAAGCAGCGACCCAGGGTGTAACTTCTGCTTATTAATAGAATG 4526  
Db 721 AlaAspArgValAlaLysGlnAlaGlnGlyValAsnLeuLeuProIleIleGluMet 740  
Qy 4527 CCCAAGCCCGACAGACCCAGACGACGTACACCTCTAGAAGACTGCGAAGACATAAAAG 4586  
Db 741 ProLysAlaProGluProArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760  
Qy 4587 ATAGACCAAGTTCTCTGAGACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAGAAATC 4646  
Db 761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 780  
Qy 4647 CTGCCCCCAAGAGGGTTAGAATATGTCACACAGATACATGCTTAACCCACTAGGA 4706  
Db 781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
Qy 4707 ACTAAACACCTCGACGCTGCTCAGACATCCCTTATCATGTTCTGAGGCTACCAAGGA 4766  
Db 801 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
Qy 4767 GTGGTCTGACTCGGTGCTCAACATTTGTGTCCTCCGACGCTGGTTAATGCTTAATCTTCC 4826  
Db 821 VallalaspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
Qy 4827 AGAATGCTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGCGCTCACTGGGAAGTGGAC 4886  
Db 841 ArgMetProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860  
Qy 4887 TTCACGTAGGTAAAGCCGCTTAATACCGGAACAAATACCTATTGGTTTTGTAGACACC 4946  
Db 861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
Qy 4947 TTTTCAGGATGGGAGAGCGCTTATCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 5006  
Db 881 PheSerGlyTyrValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900

Qy 5007 AAAATACTGGAAGAAATTTTCCAAAGATTGGAAATACCTAAGGTAATAGAGTCCAGACAAT 5066  
Db 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
Qy 5067 GGTCCACGCTTTTGTGCCAGGTAAAGTCAGGAGCTGGCCCAAGATATTGGGGAATTGATTGG 5126  
Db 921 GlyProAlaPheValAlaHisValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTyr 940  
Qy 5127 AAATGCAATTGTGCATACAGACCCCAAGCTCAGGACAGGTAGAGAGGATGAATAGAAC 5186  
Db 941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
Qy 5187 ATTAAGAGACCTTACTAAATTGACCGCGAGACTGCGCTTAATGATTGATAGCTCTC 5246  
Db 961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTrpIleAlaLeu 980  
Qy 5247 CTGCCCTTCTGTCTTTTAGGGTTAGGAACACCCCTCGACAGATTTGGGCTGACCCCTAT 5306  
Db 981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
Qy 5307 GAATTACTCTACGGGGGACCCCTCATTTGGTAGAAATTCCTTCTGTACATAGTCTGAC 5366  
Db 1001 GluLeuLeuTyrGlyGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1020  
Qy 5367 GTGCTGCTTTCCAGCCCTTTGTTCTTAGGCTCAAGGCACCTTGAGTGGGTGAGACAACGA 5426  
Db 1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
Qy 5427 GCGTGGAGGCAACTCCCGGAGGCTCTCTCAGGAGGAGGAGACTTGCAGATCCCAATCGT 5486  
Db 1041 AlaTrpArgGlnLeuArgGluAlaTyrSerGlyGlyAspLeuGlnIleProHisArg 1060  
Qy 5487 TTCCAATGGGAGATTCACTACGTTAGACGCCACCCGTCGACGAGAAACCTCGAGACTCGG 5546  
Db 1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
Qy 5547 TGGNAGGCGCTTATCTGCTACTTTTGACACACCAACCGCTGTGAAGTCGAGGAATC 5606  
Db 1081 TrpLysGlyProTyrLeuValLeuLeuThrProThrAlaValLysValGluGlyIle 1100  
Qy 5607 TCACCTCGGATCCATGCATCCACGTTAAACCGGCGCACCTCCCGATTGGGGTGGAAA 5666  
Db 1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTyrLys 1120  
Qy 5667 GCCGAAAAGACTGAAAATCCCTTAAGCTTCGCTCCATCCGCTGGTTCCTTACTCTGTC 5726  
Db 1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1140  
Qy 5727 AATAACCTCTCAGAC 5741  
Db 1141 AsnAsnLeuSerAsp 1145  
RESULT 10  
Q4VFZ2\_9GMR PRELIMINARY; PRT; 1145 AA.  
AC Q4VFZ2;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Pol protein.  
GN Name=pol;  
OS Porcine endogenous retrovirus C/A.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gammaretrovirus; 1-Mammalian type C virus group.  
OX NCBI\_TaxID=286072;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C/A 4;  
RA Denner J., Karlas A., Votteler Jr.;  
RT "Nucleotide sequence and extended characterization of a high passage  
human cell-adapted recombinant PERV-C/A.";  
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY953542; AAY28928.1; -, Genomic DNA.  
SQ SEQUENCE 1145 AA; 127925 MW; 62014B9629B04051 CRC64;

## Alignment Scores:

Pred. No.: 0 Length: 1145  
Score: 5986.00 Matches: 1134  
Percent Similarity: 99.6% Conservative: 5  
Best Local Similarity: 99.1% Mismatches: 5  
Query Match: 40.9% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3 (1-8132) x Q4VP22\_9GMR (1-1145)

```
QY 2307 ATGGGTGCGCCAGGGCAACACAGATATCCATGGACTACCCGAAACAGATTGACTTGGGA 2366
Dy 1 MetGlyAlaThrGlyGlnGlnGlnTyProTrpThrThrArgAsgThrValAspLeuGly 20
QY 2367 GTGGGCGGGTAACCCACTCTCTTCTGTCATACCTGAGTCCCGACACCCCTTTAGGT 2426
Dy 21 ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 40
QY 2427 AGAGACTTATTGACCAAGATGGGAGCACAAATTTCTTTGAAACAGGGAACACAGAAGTG 2486
Dy 41 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyLysProGluVal 60
QY 2487 TCTGCAATTAACAAACCTTATCAGTGTGTGACCTCCCAATPAGATGACGAATATCGACTA 2546
Dy 61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyArgLeu 80
QY 2547 TACTCTCCCTAGTAAGCCCTGATCAAAATATACAAATCTGGTTGGGAACAGTTTCCCAA 2606
Dy 81 TyrSerProLeuValLysProAspGlnAsnIleGlnPheTrpLeuGluGlnPheProGln 100
QY 2607 GCCTGGGCGAGAAACCGCAGGATGGGTTTGGCAAGCAAGTTCCCCACAAAGTTATTCAA 2666
Dy 101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 120
QY 2667 CTGAAGCCAGTGCACACCAAGTCTCAGTCAGACAGTACCCCTTGAGTAAGAAGCTCAA 2726
Dy 121 LeuLysAlaSerAlaThrProValSerValArgGlnTyProLeuSerLysGluAlaGln 140
QY 2727 GAAGGAATTCGGCCGCGATGCCAAAGATTAAATCAACACAGGCGATCTAGTTCCTGTCCAA 2786
Dy 141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlyIleLeuValProValGln 160
QY 2787 TCTCCCTGGAAATCTCCCTGCTACCGGTAGAAAGCTGGAGCTAATGACTATCGACCA 2846
Dy 161 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyArgPro 180
QY 2847 GTACAGACTTGAGAGAGGTCAATAAACCAGGTGCGAGGATATACACCCCAACAGTCCCGAAC 2906
Dy 181 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200
QY 2907 CTTTATAACCTCTTGTTGTCTCCACCCCAACGAGCTGGTATACAGTATTGGACTTA 2966
Dy 201 ProTyArgAsnLeuLeuCysAlaLeuProProGlnArgSerTrpTyThrValLeuAspLeu 220
QY 2967 AAGGATCCCTTCTTCGCTGAGATTACCCCACTAGCCAAACACATTTTGGCTTCGAA 3026
Dy 221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240
QY 3027 TGGAGAGATCCAGGTACGGGAAGAACCGGCGAGCTCACCTGGAGCCCGACTGCCCAAGGG 3086
Dy 241 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260
QY 3087 TTCAGAACTCCCGGACATCTTTTGGACGAAGCCCTACAGAGACCTGGCCAACTTCAGG 3146
Dy 261 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280
QY 3147 ATCCACACCTCAGGTGACCTCTCCAGTACGTGATGACCTGCTTCTGGGGGAGCC 3206
Dy 281 IleGlnHisProGlnValThrLeuLeuGlnTyValAspAspLeuLeuAlaGlyAla 300
```

```
QY 3207 ACCAAACAGGACTGCTTAGAAGGCAACGAGCAGCTACTGTGTGGAATTTGTCTGACTAGCC 3266
Dy ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 320
QY 3267 TACAGACCTCTCTAAGAGGCGCCAGATTTCAGGAGAGAGGTAAACATATTCTTGGGTAC 3326
Dy TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyLysLeuGlyTy 340
QY 3327 AGTTTGGGCGACGGGACGCGATGGCTGACGAGGACGCGAAGAAACCTGTAGTCCAGATA 3386
Dy 341 SerLeuArgAspGlyGlnAsgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 360
QY 3387 CCGGCCCAACACAGCAACAAACATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGA 3446
Dy 361 ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 380
QY 3447 CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCCACTCTACCCGCTAACCAAAAGAAAAA 3506
Dy 381 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyProLeuThrLysGluLys 400
QY 3507 GGGGAATTCTCTGGGCTCTCTGAGCACCAAGAGCATTTGTATGCTATCAAAAAGCCCTG 3566
Dy 401 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420
QY 3567 CTGAGCGCACCTCTCTGGCCCTCCCTGACGTAACTAAACCTTTACCTTTTATGTGAT 3626
Dy 421 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyValAsp 440
QY 3627 GAGGTAAAGGAGTAGCCCGGGAGTTTAAACCCMAACCTTAGGACCATGGAGAGACCT 3686
Dy 441 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 460
QY 3687 GTCCCTTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCTCAAG 3746
Dy 461 ValAlaTyLysSerLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLys 480
QY 3747 GCTATGCGAGCTGTGGCCATATCTGGTCAAGGACGCTGACAAATGACTTTGGGACAGA 3806
Dy 481 AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 500
QY 3807 ATAACTGTATAGCCCCCGATGCTATGGAGAAACATCGTTGGCAGCCCCCAGACCGATGG 3866
Dy 501 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 520
QY 3867 ATGACCAACCCCGCATGACCATCAAGCTCTCTTCACAGAGAGGCTCACGCTTC 3926
Dy 521 MetThrAsnAlaArgMetThrHisTyLysGlnSerLeuLeuLeuThrGluArgValThrPhe 540
QY 3927 GCTCCACCGAGCTCTCAACCCCTGCCACTCTTCTGCTGAAGAGACTGATGAACCAATG 3986
Dy 541 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 560
QY 3987 ACTCATGATTGCCATCACTATTGATTGAGAGACTGGGGTCCGCAAGGACCTTACAGAC 4046
Dy 561 ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 580
QY 4047 ATACCGCTGACTCGAGAAGTCTAACTGCTCTCACTGACGGAAGACAGCTATGTGGTGA 4106
Dy 581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyValValGlu 600
QY 4107 GGTAAAGAGGATGCTGGGGCGGGTGGTGGAGCGGACCCGACAGATCTGGGCCAGCAGC 4166
Dy 601 GlyLysArgMetAlaGlyAlaAlaValAspGlyThrArgThrIleTrpAlaSerSer 620
QY 4167 CTCGCGAAGGAATTCAGCACAAAAGCTGAGCTCATGGCCCTCAGCGAAGCTTTCGGG 4226
Dy 621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640
QY 4227 CTCGCGAAGGAATTCATAAATTTATACGACAGCAGAGGTATGCTTTGCGACTGCA 4286
Dy 641 LeuAlaGluGlyLysSerIleAsnIleTyThrAspSerArgTyAlaPheAlaThrAla 660
QY 4287 CACGTACATGGGCGCCATCTATAAACAAGGGGGTGTCTTACCTCAGCAGGAGGGAATA 4346
```

661 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
4347 AAGAACAAAGAGGAAATTTCTAAGCCTATTAGAGCCGTACATTTACCAAAAGCGTAGCT 4406  
681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 700  
4407 ATTATACACTCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCAAGATG 4466  
701 IleIleHisCysProGlyHisGlnLysAlaLysAspProIleSerArgGlyAsnGlnMet 720  
4467 GCTGACCGGGTTGCCAAGCAGCGACCCAGAGGTGTAACTCTCTGCGCTATATAGAAATG 4526  
721 AlaAspArgValAlaLysGlnAlaGlnGlyValAsnLeuLeuProMetIleGluThr 740  
4527 CCCAAAGCCCCAGAACCCAGACGACGATACACCTTAGAAGACTGGCAGAGATAAAAAG 4586  
741 ProLysAlaProGluProGlyArgGlnTyrThrLeuGluAspTyrGlnGluIleLysLys 760  
4587 ATAGACCAAGTTCTCTGAGACTCCGAGGAGGACCTGCTATACCTCAGATGGGAAGCAATC 4646  
761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 780  
4647 CTGCCCCACAAAGAGGGTTAGAATATGTCACACAGATACATCGTCTAACCCACCTAGGA 4706  
781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
4707 ACTAAACACTGCGAGCAGTGGTCAGAACATCCCTTATCATGTTCTGAGGCTACACGGA 4766  
801 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
4767 GTGCTGACTCGGTGGTCAACATGTGTGCTCCGACGCTGGTGTATGCTAAATCCTTCC 4826  
821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
4827 AGAATGCTCTCCAGGAAGAGACTAAGGGGAAGCACCCAGCGCTCACTGGGAAGTGGAC 4886  
841 ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTyrGluValAsp 860  
4887 TTCACTAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGGTTTTGTAGACACC 4946  
861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
4947 TTTTCAGATGGGTAGAGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 5006  
881 PheSerGlyTyrValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900  
5007 AAAATACTGGAAGAAATTTTCCAGATTTCGAATACCTAAGGTAATAGGCTCAGACAAT 5066  
901 LysIleLeuGluGluIlePheProArgPheGlyLeuProLysValIleGlySerAspAsn 920  
5067 GGTCCAGCTTTTGTGCCAGGTAAAGTCAGGACTGGCCAGATATTTGGGATTTGATGG 5126  
921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTyr 940  
5127 AAATCGCATTCGATACAGACCCCAAGCTCAGACAGGTAGAGAGTCAATAGAAC 5186  
941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
5187 ATTAAGAGACCCCTTACTAAATTCACCCGCGAGACTGGCGTGTAAATGATTGATAGCTCTC 5246  
961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTyrIleAlaLeu 980  
5247 CTGCGCTTTGTGCTTTTAGGGTTAGGAACACCCCTGGACAGTGTGGCTGACCCCTTAT 5306  
981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
5307 GAATTAATCTACGGGGGACCCCGCTTGTAGAAATTTGCTTCTGTACATAGTGTGAC 5366  
1001 GluLeuLeuTyrGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1020  
5367 GTGCTGCTTTCCAGCCCTTGTCTCTAGGCTCAAGGCATCTGAGTGGTGGTCAACGGA 5426

1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTyrValArgGlnArg 1040  
5427 GCGTGGAGGCNACTCCCGGAGGCTTACTCAGGAGGAGAGACTTGCAGATCCCATCGT 5486  
1041 AlaTyrArgGlnLeuArgGluAlaTyrSerGlyGlyAspLeuGlnIleProHisArg 1060  
5487 TTCCAAGTGGGAGATTTCAGTCTAGCTTAGAGCGCCACCGTGCAGGAAACCTCGAGACTCGG 5546  
1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
5547 TGGAAAGGCGCTTATCTCTGACTTTTGCACACACCAACCGGCTGTGAAAGTCGAAGGAATC 5606  
1081 TrpLysGlyProTyrHisValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1100  
5607 TCACACTGGATCCATGATCCACGTTAAACCGCGGCCACCTCCGATTCCGGGTGGAAA 5666  
1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTrpLys 1120  
5667 GCCGAAAAGACTGAAAATCCCTTAAAGCTTCGCTCCATCGCTGGTTCCTTACTCTGTC 5726  
1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1140  
5727 AATAACCTCTCA 5738  
1141 AsnAsnPheSer 1144  
RESULT 11  
Q8UMP5\_9GAMR  
ID Q8UMP5\_9GAMR PRELIMINARY; PRT; 1146 AA.  
AC Q8UMPS;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Pol.  
OS Porcine endogenous retrovirus.  
OC Viruses; Retrov. viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21568379; PubMed=11711616;  
DOI=10.1128/JVI.75.24.12252-12265.2001;  
Herring C., Quinn G., Bower R., Parsons N., Logan N.A., Brawley A.,  
Elsome K., Whittam A., Fernandez-Suarez X.M., Cunningham D.,  
Onions D., Langford G., Scobie L.;  
RT "Mapping full-length porcine endogenous retroviruses in a large white  
pig.";  
J. Virol. 75:12252-12265(2001).  
RL EMBL; AY056035; AAL38193.1; -, Genomic\_DNA.  
DR HSP; P03355; 116J.  
DR MEROPS; A02.020; -.  
DR DR; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:protein-dependent DNA replication; IEA.  
DR InterPro; IPR001995; Peptidase\_A2\_cat.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF000075; KnaseH; 1.  
DR Pfam; PF00065; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS00879; RNASE\_H; 1.  
SQ SEQUENCE 1146 AA; 128417 MW; 1244B29F5C270DE CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 1146

Score:	5863.00	Matches:	1109
Percent Similarity:	98.6%	Conservative:	17
Best Local Similarity:	97.1%	Mismatches:	16
Query Match:	40.1%	Indels:	0
DB:	2	Gaps:	0
US-10-723-552-3 (1-8132) x Q8UMPS_9GNMR (1-1146)			
QY	2307	ATGGGTGCCACGAGGCAACAACAGTATCCATGGACTACCCGGAAGACAGATTGACTTGGGA	2366
DB	1	MetGlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgGThrValAspLeuGly	20
QY	2367	GTGGGACGGGTAAACCCACTCTCTTGTGTCTATACCTGAGTGCCGACAGCCCTCTTAGGT	2426
DB	21	ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly	40
QY	2427	AGAGACTTATTGACCAAGATGGGAGCACAAATTTCTTTGAAACAGGGAACACAGAAGTG	2486
DB	41	ArgaspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal	60
QY	2487	TCTGCAATAACAAACCTTATCAGTGTGTGACCCCTCCAATTAGATGACGAATATCGACTA	2546
DB	61	SerValAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu	80
QY	2547	TACTCTCCCTAGTAAGCCCTGATCAAAATATACAATCTTGTTGGACAGTTTCCCAA	2606
DB	81	TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGln	100
QY	2607	GCCTGGGACAGAAACCGCAGGATCGGTTTGGCAAGCAGTTCCCCACAGTTATTCAA	2666
DB	101	AlaTrpAlaGluThrAlaGlyIleGlyLeuAlaLysGlnValProProGlnValIleGln	120
QY	2667	CTGAAGGCCAGTGCCACACAGTGTCACTGTCAGTCAGACAGTACCCCTTGAGTAAAGAGCTCAA	2726
DB	121	LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaArg	140
QY	2727	GAAGGAATTCGGCGCATGTCCAAAGATTATATCAACACAGGCGATCTAGTTCTGTCCAA	2786
DB	141	GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuAlaProValGln	160
QY	2787	TCTCCCTGGGAATCTCCCTGTACCGGTAGAAAGCTCGGACTATGACTATCCGACCA	2846
DB	161	SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro	180
QY	2847	GTACAGGACTTGAGAGAGGTCAATAAACGGGTGACGATATACACCCAAACAGTCCCGAAC	2906
DB	181	ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn	200
QY	2907	CCTTATAACCTTGTGTGTCTCCCAACCCACGAGCTGGTATACAGTATTGGACTTA	2966
DB	201	ProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeu	220
QY	2967	AAGGATCCCTTCTCTGCTGAGATTACCCCACTAGCCAAACCACTTTTGGCTTCGAA	3026
DB	221	LysAspAlaPhePheCysLeuArgGluHisProThrSerGlnProLeuPheAlaPheGlu	240
QY	3027	TGAGAGATCCAGGTACGGGAAGAACCGGCGAGCTCACCTCGACCCGACCTGCCAAGGG	3086
DB	241	TrpArgaspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly	260
QY	3087	TTCGAAGACTCCCGACCATCTTTTGCAGAGCCCTACACAGACCTGGCCCACTTCAGG	3146
DB	261	PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg	280
QY	3147	ATCCNAACCTCAGGTGACCTCTCCAGTACGTGTGATGACCTCTCTGGCGGAGGCC	3206
DB	281	IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla	300
QY	3207	ACCAAAACAGGACTGCTTAGAAGGCACGAAGCCACTACTGCTGGAAATGTCTGACCTAGGC	3266
DB	301	ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly	320
QY	3267	TACAGACCTCTCTGTAAGAAGGCCAGATTTCGAGGAGAGGATTAACATATCTGGGGTAC	3326



681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 700  
4407 ATTATACACTCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCATG 4466  
701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720  
4467 GCTCAGCGGTTGCCAGCAGCAGCCAGCGGTGTTAACTCTCTGCTCTATAATAGAAATG 4526  
721 AlaAspArgValAlaLysGlnAlaGlnAlaValAsnLeuLeuProIleIleGluThr 740  
4527 CCCAAAGCCCGAACCCAGACGACGATACACCTCAGAGACTCGCAGAGATATAAAG 4586  
741 ProLysAlaProGluProArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760  
4587 ATAGACCAAGTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATC 4646  
761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle 780  
4647 CTGCCCCACAAGAGGGTTAGAATATGTCACACAGATACATCTCTAAACCCACTAGGA 4706  
781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
4707 ACTAAACACCTGCAGCAGTGGTCAGAACATCCCTTTATCATGTTCTCAGGCTACCA 4766  
801 ThrLysHisLeuGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
4767 GTGCTGACTCGGTGGTCAACATTTGTGCTCCGCTGCCAGCTGGTTAATGCTAATCCTTC 4826  
821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
4827 AGAATGCTCCTCAGGAGAGACTAAGGGAGAGCCAGCCAGCGCTCACTGGGAAGTGGAC 4886  
841 ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860  
4887 TTCACTGAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGGTTTTGTAGACACC 4946  
861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
4947 TTTTCAGATGGGTAGAGCTTATCCTACTAAGAAAGACATCAACCGTGGTGGCTAAA 5006  
881 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900  
5007 AAAATACTGGAAGAAATTTTCCAAAGATTTCGAATACCTTAAGTAATAGGCTCAGACAAT 5066  
901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
5067 GGTCCAGCTTTTGTGCCAGGTAAAGTCAGGACTGGCCAAAGATATTGGGATTTGATTGG 5126  
921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
5127 AAATGCAATTGTGCATACAGACCCCAAGCTCAGGACAGGTAGAGAGTGAATAGAAC 5186  
941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
5187 ATTAAGAGACCTTTACTAAATTCACCGCAGACTGGCGTTAATGATTGATGATGCTC 5246  
961 IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeu 980  
5247 CTGCCCTTTGTGCTTTTATAGGTTAGGAACACCCCTCGACAGTTTGGGCTGACCCCTAT 5306  
981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
5307 GAATTACTCTACGGGGGACCCCCCACTGGTAGAATTTGCTTCTGTACATAGTGTGAC 5366  
1001 GluLeuLeuTyrGlyIleProProProLeuValGluIleAlaSerValHisSerAlaAsp 1020  
5367 GTGCTGCTTTCCAGCCTTTGCTCTAGGCTCAAGGCATTTAGTGGTGGTCAGACAACGA 5426  
1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
5427 GCGTGGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGGACTTCGACATCCCATCGT 5486  
1041 AlaTrpLysGlnLeuArgGluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArg 1060

5487 TTCCAAGTGGAGATTTCAGTCTAGCTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGG 5546  
1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
5547 TGAAGGCGCCTTATCTCGTACTTTTCCACACACCAACCGCTGTGAAAGTCGAAGGAATC 5606  
1081 TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1100  
5607 TCACCTGGATCCATGCATCCACGTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAA 5666  
1101 SerThrTrpIleHisAlaSerHisValLysLeuAlaProProAspSerGlyTrpArg 1120  
5667 GCCGAAAGACTGAAATCCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCTC 5726  
1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsn 1140  
5727 AATAAC 5732  
1141 AsnAsn 1142  
RESULT 12  
Q8UM99\_9GAMR  
ID Q8UM99\_9GAMR PRELIMINARY; PRT; 1144 AA.  
AC Q8UM99;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Pol protein.  
GN Name=pol;  
OS Porcine endogenous retrovirus.  
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20219394; PubMed=10756014;  
RX DOI=10.1128/JVI.74.9.4028-4038.2000;  
RA Czauderna F., Fischer N., Boiler K., Kurth R., Toenjes R.R.;  
RT "Establishment and characterization of molecular clones of porcine  
endogenous retroviruses replicating on human cells.";  
RL J. Virol. 74:4028-4038(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21851083; PubMed=11861838;  
RX DOI=10.1128/JVI.76.6.2714-2720.2002;  
RA Niebert M., Rogel-Gaillard C., Chardon P., Toenjes R.R.;  
RT "Characterization of Chromosomally Assigned Replication-Competent Gamma  
Porcine Endogenous Retroviruses Derived from a Large White Pig and  
Expression in Human Cells.";  
RL J. Virol. 76:2714-2720(2002).  
DR EMBL; AJ279056; CAC82502.1; -, Genomic\_DNA.  
DR HSPSP; P03355; I16J.  
DR MEROPS; A02.020; -.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR00477; RVTse.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00078; rve; 1.  
DR Pfam; PF00078; RVT 1; 1.  
DR PROSITE; PS50879; RNASE H; 1.  
SQ SEQUENCE 1144 AA; 127926 MW; 5F3AA5A5EFB98038 CRC64;  
Alignment Scores: 0 Length: 1144  
Pred. NO.: 5840.50 Matches: 1107  
Score: 97.8% Percent Similarity: 12

Best Local Similarity: 96.8% Mismatches: 24  
 Query Match: 39.9% Indels: 1  
 DB: 2 Gaps: 1

US-10-723-552-3 (1-8132) x Q8UM99\_9GAMR (1-1144)

QY 2307 ATGGGTGCGCAGGGCAACAACAGTATCCATGGACTACCCGAAAGAACAGATTGCTGGGA 2366  
 DB 1 MetGlyAlaThrGlyGlnGlnTyrProTrpThrThrArgArgThrValAspLeuGly 20  
 QY 2367 GTGGGACGGGTAAACCCACTGTTCTGTCTATACCTGAGTCCCGCAGCACCCCTCTTAGGT 2426  
 DB 21 ValGlyArgValThrGlnSerPheLeuGlyIleProGluCysProGluProLeuLeuGly 40  
 QY 2427 AGAGACTTATTACCAAGATGGGAGCACAAATTTCTTTGAACAAGGGAACACAGAAGTG 2486  
 DB 41 ArgAspLeuLeuThrThrMetGlyArgGlnIlePheGluGlnGlyThrProGluVal 60  
 QY 2487 TCTGCAATAACAACCTATCAGTGTGTGACCTCCAAATAGATGACGAATATCGACTA 2546  
 DB 61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu 80  
 QY 2547 TACTCTCCCTAGTAAAGCCGTGATCAAAATATATACAATTTCTGTTGGAAACAGTTTCCCAA 2606  
 DB 81 TyrSerSerLeuValLysProAspGlnAsnIleGlnPheTrpLeuGluGlnPheProGln 100  
 QY 2607 GCCTGGGCAGAAACCGCAGGATGGGTTTGGCAAGCAAGTTCCCCACCAAGTTATTCAA 2666  
 DB 101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 120  
 QY 2667 CTGAGCCAGTGCCACACAGTGTGAGTCAGTCAGACAGTACCCCTTGAGTAAGAGCTCAA 2726  
 DB 121 LeuLysAlaSerAlaAlaProValSerValArgGlnTyrProLeuSerLysGluAlaArg 140  
 QY 2727 GAAGGAATTCGGCGCATGTCCTCAAGATTAATCAACAGGSCATCTAGTTCCTGTCCAA 2786  
 DB 141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 160  
 QY 2787 TCTCCCTGGAAATCTCCCTGCTACCGGTAGAAAGCCTGGGACTATGACTATCGACCA 2846  
 DB 161 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 180  
 QY 2847 GTACAGACTTGAGAGAGGTCAATAAACGGGTGCGAGTATACACCAACAGTCCCGAAC 2906  
 DB 181 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200  
 QY 2907 CCTTATAACCTCTGTGTCTCTCCACCCCAACGGAGCTGGTATACAGTATTGGACTTA 2966  
 DB 201 ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTrpTyrThrValLeuAspLeu 220  
 QY 2967 AAGGATGCCCTTCTTCTGCTGAGATTACACCCCACTAGCCAAACCACTTTTGGCTTCGAA 3026  
 DB 221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240  
 QY 3027 TGGAGAGATCCAGGTACGGGAAGAACCGGCGAGCTCACTGGACCCGACTGCCCCCAAGGG 3086  
 DB 241 TrpArgAspProGlyAlaGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260  
 QY 3087 TTCAGAACCTCCCGACCATCTTTGAGAGCCCTTACAGAGACCTGGCCCACTTCAGG 3146  
 DB 261 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280  
 QY 3147 ATCCAAACACCTCAGGTGACCTCCCTCCAGTACGTGGATCACTGCTTCTGGCGGGAGCC 3206  
 DB 281 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla 300  
 QY 3207 ACCAAACAGACTGCTTTAGAAGGCAAGAAGCACTACTGCTGGAATGTCTGACCTAGGC 3266  
 DB 301 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 320  
 QY 3267 TACAGACCTCTCTTAAGAGGCCAGATTGAGAGAGAGAGTAACTACTTGGGGTAC 3326  
 DB 321 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 340

QY 3327 AGTTTGGGACGGGACGATGCTGACGAGGACGAGGACGGAAGAACTGTAGTCCAGATA 3386  
 DB SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysArgThrValValGlnIle 360  
 QY 3387 CCGGCCCAACACAGCAGCAACAAATGAGAGAGTTTTTGGGACAGCTGGATTTGCAGA 3446  
 DB ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 380  
 QY 3447 CTGTGGATCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTAAACCAAGAAAAA 3506  
 DB LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLys 400  
 QY 3507 GGGGAATTCCTCGGGCTCTGAGCACCAAGGCAATTTGATGTATCAAAAAGGCCCTG 3566  
 DB GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420  
 QY 3567 CTGAGCGCACTGCTCTGGCCCTCCCTGACGTAACTAAACCTTTACCTTTATGTGAT 3626  
 DB LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 440  
 QY 3627 GAGCGTAAGGAGTAGCCCGGGAGTTTTAAACCAACCCCTAGGACCATCGAGAAGACCT 3686  
 DB GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 460  
 QY 3687 GTCGCCTTACCTGTCAAGAAAGCTCGATCCTGTGTAGCCAGTGTGGGCCCATATGCTCAAG 3746  
 DB ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLys 480  
 QY 3747 GCTATGCGAGCTGTGGCCATACCTGCTCAAGACGCTGACAAATTTGACTTTGGGACAGAAT 3806  
 DB AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 500  
 QY 3807 ATAACGTATAGCCCCCATGATTGGAGAAACATCGTTCGGCAGCCCCCAGACCCGATGG 3866  
 DB IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 520  
 QY 3867 ATGACCAACCCCGCATGACCCACTATCAAAAGCTGCTTCTCACAGAGAGGCTCACGCTTC 3926  
 DB MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 540  
 QY 3927 GCTCCACAGCCCTCTCAACCCCTGCCACTCTTCTGCTCAAGAGACTGATGAACCAAGT 3986  
 DB AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 560  
 QY 3987 ACTCATGATTGCCATCAACTATTGATTGAGACACTGGGCTCCGCAAGGACCTTACAGAC 4046  
 DB ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 580  
 QY 4047 ATACCCGTGACTGGAGAAGTGCTAACTGTTCACTGACGGAAGAGCTATGTGGTGGAA 4106  
 DB IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValLys 600  
 QY 4107 GGTAAAGAGATGCTGGGCGGGTGTGTGGACCGGACCCGACAGTCTCGGCCACGACG 4166  
 DB GlyLysArgMetAlaGlyProProValValAspGlyThrArgThrIleTrpAlaSerSer 620  
 QY 4167 CTGCGGGAAGGAACTTCAGCACAAAAGGCTGAGCTCATGCGCCCTCAGCGAAGCTTTCGG 4226  
 DB LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
 QY 4227 CTGCGCAAGGGGAAATCCATAACATTATACCGACAGAGGTATGCTTTGGAGCTGCA 4286  
 DB LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 660  
 QY 4287 CAGTACATGGGCCCATATATAACAAAGGGGTTGCTTACCTCAGCAGGAGGGAATA 4346  
 DB HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
 QY 4347 AAGAAACAAAGAGAAATTTCTAGCCCTATTAGAGCCGTACATTTTACCAAAAAGGCTAGCT 4406  
 DB LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 700

QY 4407 ATTATACACTGCTCTGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACACGATG 4466  
DB |||||IlelleHisCysProGlyHisGlnLysAlaLysAspProIleSerArgGlyAsnGlnMet 720  
QY 4467 GCTCACCAGGTTGCCAGCAGCCAGCCAGCGGTGTTAACTCTCTGCTTATATAGAAATG 4526  
DB |||||AlaAspArgValAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProMetIleGluThr 740  
QY 4527 CCCAAAGCCCCAGAACCCAGACGACGTACACCTCTAGAGACTGGCAGAGATATAAAG 4586  
DB |||||ProLysAlaProGluProGlyArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760  
QY 4587 ATAGACCAAGTCTCTGAGACTCCGAAAGGACCTGCTATACCTCAGATGGGAAGAAATC 4646  
DB |||||IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 780  
QY 4647 CTGCCCCACAAAGAGGTTAGAAATATGTCCAACAGATACATCGTCTAACCCACTAGGA 4706  
DB |||||LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
QY 4707 ACTAAACACCTGCGACGAGTGGTTCAGAACATCCCTTATCATGTTCTGAGGCTACCGGA 4766  
DB |||||ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
QY 4767 GTGCTGACTCGGTGGTCAAACTTGTGTGCCCTGCCAGCTGGTAAATGCTAAATCCTTCC 4826  
DB |||||ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
QY 4827 AGAATGCTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGCGCTCACTGGGAAGTGGAC 4886  
DB |||||ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860  
QY 4887 TTCACCTGAGGTAAAGCCGCTAAATACGGAACAAATACCTATTGGTTTTGTAGACACC 4946  
DB |||||PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
QY 4947 TTTTCAGATGGGTAGAGGCTTATCTACTAAGAAAGAGACTCAACCGTGGTGGCTAAA 5006  
DB |||||PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900  
QY 5007 AAAATACTGGAAGAAATTTTCCAGATTTTGGAAATCACTAAGGTAAATAGGTCACACAAT 5066  
DB |||||LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
QY 5067 GCTCCAGCTTTGTCGCCAGGTAAGTCAGGACTGGCAGAGATATGGGATTTGATGG 5126  
DB |||||GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
QY 5127 AAATGCAATTCGCATACAGACCCCAAGCTCAGACAGAGTGGAGAGATGAATAGAAC 5186  
DB |||||LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
QY 5187 ATTAAGAGACCCCTTACTAAATTCACCGCGAGACTGGCGTTAATGATGGATGCTCTC 5246  
DB |||||IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeu 980  
QY 5247 CTGCCCTTTGCTTTTAGGGTTAGNACACCCCTGGACAGATTTGGCTGACCCCTTAT 5306  
DB |||||LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
QY 5307 GAATTAATCTACGGGGGACCCCCCACTGGTAGAAATTTGCTCTGTACATAGTGTGAC 5366  
DB |||||GluLeuLeuTyrGlyIleProProLeuAlaGluIleAlaPheAlaHisSerAlaAsp 1020  
QY 5367 GTGCTGCTTTCCAGCCCTTTGCTTCTAGGCTCAAGGCACTTGAGTGGGTGAGACAAGA 5426  
DB |||||ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTyrValArgGlnArg 1040  
QY 5427 GCGTGGAGGCAACTCCGGGAGCCCTACTCAGGAGGAGAGACTTCGATCCCATCGT 5486  
DB |||||AlaTrpLysGlnLeuArgGluAlaTyrSer---GlyIleAspLeuGlnValProHisArg 1059  
QY 5487 TTCCAAGTGGAGATTTCAGTCTAGCTTAGACGCCACCGTCGAGGAAACCTCGAGACTCGG 5546

DB |||||PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1079  
QY 5547 TGGAGGGCCCTTATCTCGTACTTTTGCACACACCAACGGCTGTGAAGTCGAAGGAATC 5606  
DB |||||TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1099  
QY 5607 TCCACCTGGATCCATGCCATCCAGTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAA 5666  
DB |||||ProThrTrpIleHisAlaPheHisValLysProAlaProProSerAspSerGlyTrpLys 1119  
QY 5667 GCCGAAAGACTGAAAATCCCTTAACTTTCGCTCCATCGCTGGTTCCTTACTCTGTC 5726  
DB |||||AlaGlnLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1139  
QY 5727 AATAACCTCTCA 5738  
DB |||||AsnAsnSerSer 1143  
RESULT 13  
Q8UM96\_9GAMR  
ID Q8UM96\_9GAMR PRELIMINARY; PRT; 1147 AA.  
AC Q8UM96;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Pol protein.  
GN Name=pol;  
OS Porcine endogenous retrovirus.  
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20219394; PubMed=10756014;  
RA DOI=10.1128/JVI.74.9.4028-4038.2000;  
RA Crauderna F., Fischer N., Boller K., Kurth R., Toenjes R.R.;  
RT "Establishment and characterization of molecular clones of porcine  
endogenous retroviruses replicating on human cells."  
RL J. Virol. 74:4028-4038 (2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21851083; PubMed=11861838;  
RX DOI=10.1128/JVI.76.6.2714-2720.2002;  
RA Niebert M., Rogel-Gaillard C., Chardon P., Toenjes R.R.;  
RT "Characterization of Chromosomally Assigned Replication-Competent Gamma  
Porcine Endogenous Retroviruses Derived from a Large White Pig and  
Expression in Human Cells."  
RL J. Virol. 76:2714-2720 (2002).  
DR EMBL; AJ279057; CAC82505.2; -; Genomic\_DNA.  
DR MEROPS; A02.020; -.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR001995; Peptidase\_A2\_cat.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR00477; RVTse.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR PROSITE; PS50175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS50879; RNASE\_H\_1.  
SQ SEQUENCE 1147 AA; 128404 MW; C6374B0CB460901D CRC64;  
Alignment Scores: 0 Length: 1147  
Pred. No.: 0

Score:	5831.00	Matches:	1104
Percent Similarity:	98.4%	Conservative:	20
Best Local Similarity:	96.7%	Mismatches:	18
Query Match:	39.8%	Indels:	0
DB:	2	Gaps:	0

US-10-723-552-3 (1-8132) x Q8UM96 9GAMR (1-1147)

Qy	2307	ATGGGTGCCACAGGCGCAACAACAGTATTCATCGACTACCCGAGAACACAGTTGACTTGGGA	2366
Db	1	MetGlyAlaThrGlyGlnArgIntyPrProTrpThrThrArgArgThrValAspLeuVal	20
Qy	2367	GTGGGACGGGTAAACCCACTCGTTTCGTGTACATCGTAGTGCCGACACCCCTCTTAGGT	2426
Db	21	ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly	40
Qy	2427	AGAGACTTATTGACCAAGATGGGAGCACAAATTTCTTTTGAAACAAGGAAACACCAAGTGT	2486
Db	41	ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal	60
Qy	2487	TCTGCAATAACAAACCTATCACTGTGTGTGACCTCCAAATAGATGACGAATATCGACTA	2546
Db	61	SerValAspLeuLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyArgLeu	80
Qy	2547	TACTCTCCCTAGTAAAGCCTGTACAAATATACAATTCGTGTTGGACACAGTTTCCCCAA	2606
Db	81	TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGln	100
Qy	2607	GCCTGGCGCAGAAACCGCAGGAGTGGTGTGGCAAAAGCAAGTTCCCTCCACAAAGTTATTCAA	2666
Db	101	AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProGlnValIleGln	120
Qy	2667	CTGAAGCCAGTGCCACACCGAGTGTCTAGTCAAGCAGTACCCCTTGTAAGAAGCTCAA	2726
Db	121	LeuLysAlaSerAlaThrProValSerValArgIntyPrProLeuSerArgGluAlaArg	140
Qy	2727	GAAGCAATTGGCCGCGCATGTCCAAGATTAAATCCAACAGGCGATCTTAGTTCTCGTCCAA	2786
Db	141	GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln	160
Qy	2787	TCTCCCTGGAAATCTCCCTCTGCTACCGGTTAGAAAAGCCTGGACTAATAGCTATCGACCA	2846
Db	161	SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro	180
Qy	2847	GTACAGACTTGAGAGAGGTCAATAACGGGTGCAGGATATACACCCAAACAGTCCCGAAC	2906
Db	181	ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn	200
Qy	2907	CTTTATAACCTCTTGTTGTCTCTCCACCCCAACGGAGCTGGTATACAGTATTGGACTTA	2966
Db	201	ProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeu	220
Qy	2967	AAGATGCTTCTTCTGCTGAGATTACCCCACTAGCCAACACATTTTGTGCTTCGAA	3026
Db	221	LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu	240
Qy	3027	TGGAGATCCAGGTACGGAGAACCGGCGAGTCACTCGGACCCGACTGCCCCCAAGG	3086
Db	241	TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly	260
Qy	3087	TTCAAGAACTCCCGACCATCTTTTGACGAAGCCCTACACAGACCTGGGCCAACTTCAGG	3146
Db	261	PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg	280
Qy	3147	ATCCAACCCCTGAGGTGACCTCTCTCAGTACGTGATGACCTGCTTCTGGCGGAGGCC	3206
Db	281	IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla	300
Qy	3207	ACCAACACGAGACTGCTTAGAAGGCACCAAGGCATCTGCTGGAAATGTCTGACCTAGGC	3266
Db	301	ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuLeuGlnLeuSerAspLeuGly	320
Qy	3267	TACAGAGCCTCTGCTAAGAAAGGCCAGATTTCGAGGAGAGAGGTAAACATACTTGGGGTAC	3326

Db 681 LysAsnLysAspGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 700  
Qy 4407 ATTATACACTGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCATG 4466  
Db 701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720  
Qy 4467 GCTCAGCGGTTGCCAAGCAGCGAGCCAGGGTGTAAACCTTCTGCTTATATAAGAAATG 4536  
Db 721 AlaAspArgIleAlaLysGlnAlaGlnAlaValAsnLeuLeuProIleIleGluThr 740  
Qy 4527 CCCAAAGCCCGAGAACCCAGACGACGATACACCCCTAGAGACTGGCAAGAGATAAAAG 4586  
Db 741 ProLysAlaProGluProArgArgGlnThrLeuGluAspTrpGlnGluIleLysLys 760  
Qy 4587 ATAGACACAGTCTCTGAGACTCCGGAAGGAGCTGTATACCTCAGATGGGAAGAAATC 4646  
Db 761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle 780  
Qy 4647 CTGCCCCACAAAGAGGTTAGATATGTCACACAGATACATGCTCTAAACCCACTAGGA 4706  
Db 781 LeuProHisLysGluGlyLeuGluThrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
Qy 4707 ACTAAACACCTGCAGCAGTGTGGTCAGAACATCCCTTATCATGTTCTGAGGCTACACGGA 4766  
Db 801 ThrLysHisLeu\*\*\*GlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
Qy 4767 GTGCTGACTCGGTGGTCAAAATGTTGTCCTGCCAGCTGGTAAATGCTAAATCCTTCC 4826  
Db 821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
Qy 4827 AGAATGCTCCAGGAGAGAGACTAAGGGAGCCACCCAGCGCTCACTGGGAAGTGGAC 4886  
Db 841 LysValProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860  
Qy 4887 TTCACTGAGGTAAAGCCGGCTAAATACAGAAACAAATACCTATTGGTTTTGTAGACACC 4946  
Db 861 PheThrLysValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
Qy 4947 TTTTCAGATGGGTAGAGCTTATCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 5006  
Db 881 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900  
Qy 5007 AAAATACTGGAAGAAATTTTCCAAAGATTTCGAATACCTAAGGTAAATAGGCTCAGACAAT 5066  
Db 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
Qy 5067 GGTCACAGCTTTGTGTCCAGGTAAGTCAGGAGCTGCCAAGATATTGGGATTGATTGG 5126  
Db 921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
Qy 5127 AAATGCAATTGTGCATACAGACCCCAAGCTCAGGACGAGTAGAGAGATGAATAGAAC 5186  
Db 941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
Qy 5187 ATTAAGAGACCCCTACTATAATTCACCGCGAGACTGCGGTTAATGATTGATAGCTCTC 5246  
Db 961 IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeu 980  
Qy 5247 CTGCCCTTTGTGCTTTTGTAGGGTTAGGAACACCCCTGACAGTGTGGGCTGACCCCTAT 5306  
Db 981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
Qy 5307 GAATTACTCTACGGGGAGACCCCCCATTTGTTAGAAATGCTTCTGTACATAGTGTGAC 5366  
Db 1001 GluLeuLeuTyrGlyGlyProProLeuValGluIleAlaProValHisSerAlaAsp 1020  
Qy 5367 GTGCTGCTTTCCAGCCTTCTCTAGGCTCAAGGCACCTTGAGTGGTGGTCAGACAAGA 5426  
Db 1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluThrPvalArgGlnArg 1040  
Qy 5427 GCGTGGAGGCAACTCCGGGAGCGCTTACTCAGGAGGAGGAGACTTGCAGATCCCATCGT 5486  
Db 1041 AlaTrpLysGlnLeuArgGluAlaTyr-serGlyGlyAspLeuGlnValProHisArg 1060

Qy 5487 TTCCAAGTGGAGATTTCAGTCTAGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGG 5546  
Db 1061 PheGlnValGlyAspSerValTyrValArgHisArgAlaGlyAsnLeuGluThrArg 1080  
Qy 5547 TGAAGGGCCCTTATCTCGTACTTTTACCACACACCAACGGCTGTGAAAGTCGAAGGAATC 5606  
Db 1081 TrpLysGlyProTyrLeuValLeuThrThrProThrAlaValLysValGluGlyIle 1100  
Qy 5607 TCACACTGGATGCCATGCCAGTTAAACCGGCGCCACCTCCCGATTTCGGGTGGAAA 5666  
Db 1101 SerThrTrpIleHisAlaSerHisValLysLeuAlaProProAspSerGlyTrpArg 1120  
Qy 5667 GCCGAAAGACTGAAATCCCTTAAGCTTCGCTCCATCGCGNGTTCCTTACTCTGTC 5726  
Db 1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgLeuValProTyr-SerAsn 1140  
Qy 5727 AATAAC 5732  
Db 1141 AsnAsn 1142  
RESULT 14  
Q8Q6U4\_9GAMR  
ID Q8Q6U4\_9GAMR PRELIMINARY; PRT; 1142 AA.  
AC Q8Q6U4;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Pol protein.  
GN Name=pol;  
OS Porcine endogenous retrovirus.  
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
ON NCBI\_TaxID=61673;  
RX NUCLEOTIDE SEQUENCE.  
RX DOI=10.1128/JVI.76.6.2714-2720.2002;  
RA Niebert M., Rogel-Gaillard C., Chardon P., Toenjes R.R.;  
RT "Characterization of Chromosomally Assigned Replication-Competent Gamma  
RT Porcine Endogenous Retroviruses Derived from a Large White Pig and  
RT Expression in Human Cells.";  
RL J. Virol. 76:2714-2720(2002).  
DR EMBL; AF435967; AAL87856.1; -; Genomic\_DNA.  
DR HSP; P03355; IMWL.  
DR MEROPS; A02.020; -.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR00477; RVTse.  
DR Pfam; PF00075; RnaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR PROSITE; PS00879; RNASE H; 1.  
SQ SEQUENCE 1142 AA; 127712 MW; B0F55CA742AAC98F CRC64;  
Alignment Scores:  
Pred. No.: 6.19e-314 Length: 1142  
Score: 5466.50 Matches: 1042  
Percent Similarity: 93.4% Conservative: 26  
Best Local Similarity: 91.1% Mismatches: 73  
Query Match: 37.3% Indels: 3  
DB: 2 Gaps: 2  
US-10-723-552-3 (1-8132) x Q8Q6U4\_9GAMR (1-1142)  
Qy 2307 ATGGTGCCACGAGCGCAACACAGTATCCATGGACTACCCGAGAACAGTTGACTTGGGA 2366  
|||||

Db	1	MetGlyAlaThrGlyGlnGlnTyrProThrThrThrArgArgThrValAspLeuGly	20
Qy	2367	GTGGACGGTAACCACTCGTTTCTGTGTCATACCTGAGTGCCCGACGACCCCTCTTAGGT	2426
Db	21	ValGlyArgValThrGlnSerPheLeuGlyIleProGluCysProGluProLeuLeuGly	40
Qy	2427	AGAGACTTATTGACCAAGATGGGAGCACAAATTTCTTTGAAACGAGGGAACCGAAGTG	2486
Db	41	ArgAspLeuLeuThrThrMetGlyArgGlnIlePhePheGluGlnGlyThrProGluVal	60
Qy	2487	TCTGCAATAACAAACCTTACCTGTTGTGACCTTCCAAATAGATGACGAATATCGACTA	2546
Db	61	SerAlaIasnAnLysProIleThrValLeuThrLeuGlnLeuAspGluTyrArgLeu	80
Qy	2547	TACTCTCCCTAGTAAGCCGTGATCAAAATATACAATTTCTGGTTGGAAACAGTTTCCCAA	2606
Db	81	PheSerProValLysLeuAspGlnAsnIleGlnPheGlySerThrGlnPheProGln	100
Qy	2607	GCCTGGGCAGAAACCGCAGGATGGTTTGGCAAGCAGTTCCTCCCAAGATTATTCAA	2666
Db	101	AlaLeuAlaGluProAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln	120
Qy	2667	CTGAAGGCCAGTGCACACCACTGTCTAGTCAGACAGTACCCTTTGAGTAAAGAAAGCTCAA	2726
Db	121	LeuLysProSerLeuAlaProValProValArgGlnSerProPheSerLysGluAlaArg	140
Qy	2727	GAAGGAATTCGGCCGATGTCCAAAGATTATCAACAGGCGATCTCTAGTTCTGTCCAA	2786
Db	141	GluGlyIleArgProHisIleValGlnArgLeuIleGlnGlnIleIleValProValGln	160
Qy	2787	TCTCTCTGGAAATCTCCCTGCTACCGGTTAGAAAGCTGGGACTAATGACTATCGACCA	2846
Db	161	SerProThrAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro	180
Qy	2847	GTACAGACTTTCAGAGAGTCAATAACGGGTGAGGATATACACCAACAGTCCCGAAC	2906
Db	181	ValGlnAspPheGluArgGlyGlnLysArgValGlnAspIleHisProThrValProAsn	200
Qy	2907	CTTTATAACCTCTTGTGTCTCTCCCAACCGAGCTGGTATACAGATTGGACTTA	2966
Db	201	ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTyrThrValLeuAspLeu	220
Qy	2967	AAGATGCTTCTTCTGCTGAGATTACACCCCACTAGCCAAACACATTTTGGCTTCGAA	3026
Db	221	LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu	240
Qy	3027	TGGAGAGATCCAGTACGGGAAGAACCGGCAGCTCACCTGGACCCGACTGCCCCAGGG	3086
Db	241	TrpArgAspProGlyAlaGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly	260
Qy	3087	TTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGG	3146
Db	261	PheLysAsnPheProThrIlePheAspGlnAlaLeuHisArgAspLeuAlaAsnPheArg	280
Qy	3147	ATCCAAACCTCAGGTGACCTCTCCAGTACGTGATGACCTGCTTTCGGCGGAGCC	3206
Db	281	IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuValLeuAlaAla	300
Qy	3207	ACCAACAGAGACTCTTAGAAGGCACCAAGSCACTACTGCTGGAAATGCTGACCTAGGC	3266
Db	301	ThrLysGlnAspCysLeuGlnArgProLysGlyLeuLeuValGlnLeuSerAspLeuGly	320
Qy	3267	TACAGACCTCTCTGAAGAGCCAGATTGTCAGAGAGAGGTAAATACTTTGGGGTAC	3326
Db	321	AspArgAlaPheGlyTyrLysAlaHisIleCysProThrGluValThrTyrLeuGlyTyr	340
Qy	3327	AGTTTGGGACGGGCGGATGCTGACGGAGCACGGAAGAAACCTGTAGTCCAGATA	3386
Db	341	ArgLeuArgGlyArgHisArgTyrLeuThrGluAlaProGlnThrThrValValGlnIle	360
Qy	3387	CCGCCCCACACAGCAACAAATGAGAGATTTTGGGACAGCTGATTTTGCAGA	3446
Db	361	ProGlyProThrProAlaLysGlnValArgGluPheLeuGlyThrValGlyPheCysArg	380
Qy	3447	CTGTGGATCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTAACCAAAAGAAAA	3506
Db	381	LeuTrpIleProGlyPheAlaThrLeuProAlaProLeuTyrProLeuProLysGluLys	400
Qy	3507	GGGGAATTCCTGGGCTCTCGAGCACAGAGGCAATTTGATCTATCAAAAGCCCTG	3566
Db	401	GlyGluPheSerTrpAlaLeuGlnHisGlnLysAlaPheAspAlaIleLysLysAlaLeu	420
Qy	3567	CTGAGCCACCTGCTCTGGCCCTCCCTGACGTACTAAACCCCTTTACCTTTATGTGAT	3626
Db	421	LeuSerAlaProAlaLeuAlaLeuProAspValThrLys-----ThrLeuTyrValAsp	438
Qy	3627	GAGCGTAAGGAGTAGCCGGGAGTTTAAACCCAAACCCCTAGGACCATGAGAGAGCCT	3686
Db	439	GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro	458
Qy	3687	GTGCCCTACCTGTCAAGAAAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCTGAAG	3746
Db	459	ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLys	478
Qy	3747	GCTATCCAGCTGTGGCCTACTGCTCAAGCAGCTGACAAATTTGACTTTGGGACAGAT	3806
Db	479	AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn	498
Qy	3807	ATAACTGTAAATAGCCCCCATGCAATTGGAGAACATCTGTCGGCAGCCCCCAGACCGATGG	3866
Db	499	IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp	518
Qy	3867	ATGACCAACCCCGCATGACCCATATCAAGCTGCTTCTCAGAGAGGGTCACTTC	3926
Db	519	MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe	538
Qy	3927	GCTCCACCGCCCTCTCAACCTGCGACCTCTCTGCTCAAGAGAGCTGACCAAGT	3986
Db	539	AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProLeu	558
Qy	3987	ACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGAC	4046
Db	559	ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp	578
Qy	4047	ATACCGCTGACTGGAGAAGTGTAACTGTTTCACTGACGGAAGCAGCTATGTGGTGA	4106
Db	579	IleProLeuThrGlyGluProValThrTrpPheThrAspGlySerSerTyrLeuValGlu	598
Qy	4107	GGTAAGAGGATGCTGGGGCGGTGGTCGACGGGACCCGACGATCTGGCCAGCAGC	4166
Db	599	GlyAsnLysMetAlaGlyAlaAlaValValAspArgThrProThrIleTrpGlyThrAsn	618
Qy	4167	CTGCCGGAAGAACTTTCAGCAAAAAGCGCTGAGCTCATGGCCCTCACGCAAGCTTTGGG	4226
Db	619	LeuProGluArgThrSerSerGlnLysGlyGluLeuIleGlyLeuMetGlnAlaPheArg	638
Qy	4227	CTGCCGGAAGGAAATCCATAAATTTATCCGACAGCAGGATATGCTTTGGACTGCA	4286
Db	639	LeuGlyGlnGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla	658
Qy	4287	CAGTCATCTGGGCGCATCTATAACAAAGGGGTGCTTACCTCAGCAGGAGGGAATA	4346
Db	659	HisValHisGlyAlaIleTyrThrGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle	678
Qy	4347	AAGAAACAAAGAGAAATTTCTAAGCCCTATTAGAACCGCTACATTTACCAAAAGCGTAGCT	4406
Db	679	LysAsnLysGluGluLeuLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla	698
Qy	4407	ATTATACATGCTCTCGACATCAAGAAAGCTTAAAGATCTCATATCCAGAGAAACAGATG	4466
Db	699	IleIleHisCysProGlyHisGlnLysAlaLysAspProIleSerArgGlyAsnGlnMet	718
Qy	4467	GCTGACCGGGTTGCCAAGCAGGCGCCCGGGTGTAACTTCTGCTTAAATAGAAATG	4526
Db	719	AlaAspArgValAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProMetIleGluThr	738

```
Qy 4527 CCCAAGCCCCAGAACCCAGACGACAGTACACCCCTAGAGACTGGCAGAGATATAAAAG 4586
Db 739 ProlysAlaProGluProGlyArgGlnThrLeuGluAspTrpGlnGluLeuLeuLeuLeu 758
Qy 4587 ATAGACCAAGTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATC 4646
Db 759 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLeuLeuLeu 778
Qy 4647 CTGCCCAACCAAGAGGGTTAGATATGTCACAGATACATCGTCTAACCCACTAGGA 4706
Db 779 LeuProHisLeuGluGlyLeuGluThrValGlnGlnIleHisArgLeuThrHisLeuGly 798
Qy 4707 ACTAAACACCTGCAGCAGCTTCGCTCAGAACATCCCTTATCATGTTCTGAGGCTACCA 4766
Db 799 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 818
Qy 4767 GTGGCTGACTCGGTGGTCAAAATTTGTGTGCGCTGCCAGCTGGTTAATGCTTAATCCT 4826
Db 819 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 838
Qy 4827 AGAATGCTCCAGGAGAGACTAAGGGAGAGCCACCCAGGCGCTCACTGGGAAGTGGAC 4886
Db 839 ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 858
Qy 4887 TTCACTGAGGTAAAGCGGCTAAATACGCAACAAATACCTATTGTTGTTGTAGACACC 4946
Db 859 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 878
Qy 4947 TTTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 5006
Db 879 PheSerGlyTrpValGluAlaTyrProThrLysLysGlnThrSerThrValValAlaLys 898
Qy 5007 AAATACTGGAAGAAATTTTCCAGATTGTCAGATACCTAAGTAAATAGGTGACAGCAAT 5066
Db 899 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 918
Qy 5067 GGTCAGCTTTGTTGCCAGGTAAAGTCCAGGACTGGCCAGATATTGGGATTGATTGG 5126
Db 919 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 938
Qy 5127 AAATGCTATTGTCATACAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAAC 5186
Db 939 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 958
Qy 5187 ATTAAAGAGACCTTACTAAATTCACCGGAGACTGCGTAAATGATGATGATGATCTC 5246
Db 959 IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeu 978
Qy 5247 CTGCGCTTTGCTTTTGGGTTAGGAAACCCCTGACAGATTTGGGCTGACCCCTAT 5306
Db 979 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 998
Qy 5307 GAATTACTCTACGGGGGACCCCCCAATTTGGTAGAAATTTGCTCTGTACATAGTGTGAC 5366
Db 999 GluLeuLeuTyrGlyGlyProProLeuAlaGluIleAlaPheAlaHisSerAlaAsp 1018
Qy 5367 GTGCTGCTTTCCAGCCTTTGTTCTTAGGCTCAAGGCACTTAGTGCTGAGACACGA 5426
Db 1019 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1038
Qy 5427 GCGTGGAGCAACTCCGGGAGCGCTACTCAGGAGGAGAGACTTCGAGATCCCATCGT 5486
Db 1039 AlaTrpLysGlnLeuArgGluAlaTyrSer---GlyGlyAspLeuGlnValProHisArg 1057
Qy 5487 TTCCAAGTGGAGATTACAGTCTAGCTTAGACGCCACCGTGCAGGAAACCTCGAGACTGG 5546
Db 1058 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1077
Qy 5547 TGGAGGGCCCTTATCTGCTACTTTTACACACCAACCGCTGTCAAGTGCAGAGGATC 5606
Db 1078 TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyLeu 1097
Qy 5607 TCCACCTGGATCCATCCACCGTTAAACCGGCGCCACCTCCCGATTTCGGGGTGGAAA 5666
```

```
Db 1098 ProThrTrpIleHisAlaPheHisValLysProAlaProProSerAspSerGlyTrpLys 1117
Qy 5667 GCCGAAAAGACTGAAATCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTC 5726
Db 1118 AlaGlnLysThrGluAsnProLeuLysLeuArgLeuHisArgValProTyrSerVal 1137
Qy 5727 AATAACCTCTCA 5738
Db 1138 AsnAsnSerSer 1141
RESULT 15
ID Q8Q6U7_9GAMR PRELIMINARY; PRT; 1139 AA.
AC Q8Q6U7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pol protein.
GN Name=pol;
OS Porcine endogenous retrovirus.
OC Viruses; Retrod viruses; Retroviridae; Gammaretrovirus;
OC 1-Mammalian type C virus group.
OX NCBI_TaxID=61673;
RN [1]
RP NUCLEOTIDE SEQUENCE
RX MEDLINE=21851083; PubMed=11861838;
RX DOI=10.1128/JVI.76.6.2714-2720.2002;
RA Niebert M., Rogel-Gaillard C., Chardon P., Toenjes R.R.;
RT "Characterization of Chromosomally Assigned Replication-Competent Gamma
RT Porcine Endogenous Retroviruses Derived from a Large White Pig and
RT Expression in Human Cells";
RL J. Virol. 76:2714-2720(2002).
DR EMBL; AF435966; AAL87853.1; -; Genomic_DNA.
DR HSP; P03355; IMML.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006313; P:DNA transposition; IEA.
DR GO; GO:0006278; P:DNA-dependent DNA replication; IEA.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00078; RVT; 1; 1.
DR PROSITE; PS50879; RNASE_H; 1.
SQ SEQUENCE 1139 AA; 127325 MW; 3B63C42301C04FAC CRC64;
Alignment Scores:
Pred. No.: 4.68e-309 Length: 1139
Score: 5383.50 Matches: 1054
Percent Similarity: 94.0% Conservative: 19
Best Local Similarity: 92.3% Mismatches: 52
Query Match: 36.8% Indels: 17
DB: 2 Gaps: 13
US-10-723-552-3 (1-8132) x Q8Q6U7_9GAMR (1-1139)
Qy 2347 GAAGAACAGTTGACTGGGAGTGGGACGGGTAAACCCACTCGTTCTGTCATACCTG--- 2403
Db 4 GluAspGlnLeuThrTrpGluLeuAspGlyGluProThrGlyPheTrpValIleLeuLys 23
Qy 2404 AGTCCAGACACCCCTCTTAGGTAGACACTATTATGACACAGATGGGAGCACAATTTTC- 2461
Db 24 Cys-ProSerThrLeuLeuArgValGlnThrTyrAspGlnAspGlyLysProLysPheLe 43
Qy 2462 -TTTGAACAGGGAACCC---AGAAGTGTCTGCAATAAACAACCTATCATCTGTGTGA 2517
```



Db 43 uPheGluProArgGluThrGlnLysCysLeu-GlnAsnThrAsnProIleThrValLeuT 63  
Qy 2518 CCCTCCCAATTAGATGACGAATATCGACTATATCTCTCCCTAGTAAAGCCTGATCAAAATA 2577  
Db 63 hrLeuGlnLeu- --AspGluTyrArgLeuTyrSerProLeuValLysProAspGlnAsnI 82  
Qy 2578 TACAATTCTGGTTGGACAGTTTCCCAAGCCTGGCGAAGAACCGCAGGGATGGTTGG 2637  
Db 82 leGlnPheTrpLeuGluGlnPheProLysAlaTrpAlaGluThrAlaGlyMetGlyLeuA 102  
Qy 2638 CAAAGCAAGTTCCCCCAAGCTTATCAACTGAAGCCAGTGCACACACCAGTGTCACTCA 2697  
Db 102 laLysGlnValProProGlnValIleGlnLeuLysAlaSerAlaAlaProValSerValA 122  
Qy 2698 GACAGTACCCCTTCAGTAAAGAGCTCAAGAGGAATTCGGCCGCATGTCCAAAGATTAA 2757  
Db 122 rgGlnTyrLeuLeuSerLysGluAlaArgGluGlyIleGlyProHisValGlnArgLeuI 142  
Qy 2758 TCCAAACAGGGCATCTAGTCTCTGTCCTCAATCTCCCTGGAATPACTCCCTGCTACCGGTTA 2817  
Db 142 leGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLeuProValA 162  
Qy 2818 GAAAGCCTGGGACTAATGACTATACGACAGTACAGGACTTCAGAGAGGTCAATAAAGCGG 2877  
Db 162 rgLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsnLysArgV 182  
Qy 2878 TGCAGGATATACACCAACAGTCCCGAACCCTTATAACCTCTGTGTGCTCTCCACCC 2937  
Db 182 alGlnAspIleHisProThrValProAsnProTyrAsnLeuLeuCysAlaLeuProProG 202  
Qy 2938 AACGGAGCTGTATACAGTATTGACTTAAAGATGCTTCTTCTGCTGAGATTACACC 2997  
Db 202 lnArgSerTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisP 222  
Qy 2998 CCCTAGCCCAACCACTTTTGGCTTCGAATGGAGAGATCCAGGTACCGGAGAACCGGGC 3057  
Db 222 roThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyAlaGlyArgThrGlyG 242  
Qy 3058 AGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTTGACCAAG 3117  
Db 242 lnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePheAspGluA 262  
Qy 3118 CCCTACACAGACCTGGCCACTTCAGGATCGAACACCTTCAGGTGACCCCTCTCCAGT 3177  
Db 262 laLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnT 282  
Qy 3178 ACGTGGATGACCTCTCTCGCGGAGCCACCAACAGGACTGCTTAGAAGGCACGAGG 3237  
Db 282 yrValAspAspLeuLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGluGly----- 299  
Qy 3238 CACTACTGCTGGAATTGCTGACCTAGGCTACAGAGCTCTGCTAAGAAGGCCAGATT 3297  
Db 300 --LeuLeuLeuGluLeuPheAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleC 319  
Qy 3298 GCAGGAGAGGTAACTACTTTGGGGTACAGTTTG---CGGGACGGGACGCA---TGGC 3351  
Db 319 ysArgArgGluAlaThrGlnLeuGlyValGlnValCysGlyAlaGlyGlnSerAspTrpL 339  
Qy 3352 TGAGC---GAGGCACGGAAGAACTGTA---GTCCAGATACCGGCCCAACACACGCCA 3405  
Db 339 eutrGlyLysAlaArgLysLysThrValGlnProLysIleGlyProProThrThrAlaL 359  
Qy 3406 AACAA---ATCAGAGAGTGTGGGGACA---GCTGGATTTCAGACTGTGGATCCCGG 3459  
Db 359 ysGlnValValArgGluPhePheGlyAlaGlnValGlyPheCysArgLeuTrpIleProG 379  
Qy 3460 GGTTCGGACCTTAGCAGCCCACTTACCCGCTTAACCAAGAAAGGGAATTCTCCT 3519  
Db 379 lyPheAlaThrLeuAlaAProLeuTyrProLeuThrLysGluLysGlyGluPheSerT 399  
Qy 3520 GGCTCTGAGCACCAGAGGCATTTCATGCTATCAAAAGGCCCTGCTAGCGCACCTG 3579  
Db 399 rpAlaLeuGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuSerSerAlaProA 419

Qy 3580 CTCTGGCCCTCCCTGACGTAAACCCCTTTTACCTTTATGTGTGATGAGCTGAAGGAG 3639  
Db 419 laLeuAlaLeuProAspValLeuLysProPheThrLeuTyrValAspGluArgLysGlyV 439  
Qy 3640 TAGCCCGGGAGTGTAAACCCAAACCCCTAGGACCATGGAGAACCTGTGCTCCCTACCTGT 3699  
Db 439 alAlaArgGlyValLeuThrGlnIleLeuGlyProTrpArgArgProValAlaTyrLeuS 459  
Qy 3700 CAAAGAAGCTCGATCTGTAGCCAGTGTGGCCCATATGCTGAAGCTATCGCAGCTG 3759  
Db 459 erLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLysAlaIleAlaVal 479  
Qy 3760 TGGCCATPACTGGTCAAGGACCTGCACAAATTCAGCTTTGGGACAGATATTAACCTGTAATAG 3819  
Db 479 alAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleA 499  
Qy 3820 CCCCCATGATGGGAGAACATGTTCCGGCAGCCGCCAGACCCGATGATGATGACCAACGCC 3879  
Db 499 laProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrpMetThrAsnAlaA 519  
Qy 3880 GCATGACCCACTATCAAGGCTGCTCTCACAGAGAGGTCACGTTCCGCTCCACACCGC 3939  
Db 519 rgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPheAlaProProAlaA 539  
Qy 3940 CTCTCAACCTGCGCCTCTCTGCTGAAGAGACTGATGAACCACTGACTCATGATTGCC 3999  
Db 539 laLeuAsnProAlaThrLeuLeuProGluThrAspGluProValThrHisAspCysH 559  
Qy 4000 ATCAACTATTGATTGAGGAGACTGGGTCCGCAAGGACCTTACAGACATA---CCGCTGA 4056  
Db 559 is---LeuLeuIleGluGluThrGlyValArgLysAspLeuThrAspIleProProLeuT 578  
Qy 4057 CTGAGAGAGTCAACCTGCTTCACTACAGGAGACGATGTGTTGGAGAGGTAAAGGA 4116  
Db 578 hrGlyLysMetLeuThrTrpPheThrAspGlySerSerTyrValValGluGlyLysSerM 598  
Qy 4117 TGGCTGGGGCGGGTGGTGGACGGGACCCGCGACGATCTGGCCGAGCAGCTTCGCGGAAG 4176  
Db 598 etAlaGlyProProValValThrGlyThrArgThrIleTrpAlaSerSerLeuProGluG 618  
Qy 4177 GAATTCAGACACAAAAGGCTGAGCTCATGCTCCCTCAGCAAGCTTTCCGGCTGCGCCGAAG 4236  
Db 618 lyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluG 638  
Qy 4237 GGAATTCATAACATTTATACGACAGCAGGTATGCTTTTGGACTGACACGCTACATG 4296  
Db 638 lyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisG 658  
Qy 4297 GGCCCATCTATAAACAAAGGGGTGCTTACCTCAGCAGGAGGGAATAAAGAACAAAG 4356  
Db 658 lyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysThrLysG 678  
Qy 4357 AGGAATTCATAACCTATTAGAACGCGTACATTTACAAAGGCTAGCTATTATACACT 4416  
Db 678 luGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIleIleHisC 698  
Qy 4417 GTCCTGACATCAGAAGCTAAAGATCTCATATCCAGAGGAACACAGATGCTGACCGGG 4476  
Db 698 ysProGlyHisGlnLysAlaLysAspProIleSerArgGlyAsnGlnMecAlaAspArgV 718  
Qy 4477 TTCCCAAGCAGGCGCCCGGGTGTAAACCTTCTGCTTATAATAGATAATGCCAAAGCCC 4536  
Db 718 alAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProMetIleGluThrProLysAlaP 738  
Qy 4537 CAGAACCCAGACGACAGTACACCTTAGAGACTGGCAAGAGATATAAAGATAGACCACT 4596  
Db 738 roGluProGlyArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnP 758  
Qy 4597 TCCTCAGACTCCGGAGGAGCCTGCTATACCTCAGATGGAGAGGAATCTCTCCCCACA 4656  
Db 758 heSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIleLeuProHisL 778



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: February 14, 2006, 12:52:14 ; Search time 216.36 Seconds  
(without alignments)  
3302.856 Million cell updates/sec

Title: US-10-723-552-3  
Perfect score: 14636  
Sequence: 1 GCGTGTGTACACTGTGGG.....CTGTTTGATCAAAAAAAA 8132

Scoring table:  
BLOSUM62 Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2443163 seqs, 439378781 residues  
Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSSWEB spool/US10723552/runat\_14022006\_125139\_12833/app\_query.fasta\_1  
-DB=A Genesep -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abes02p  
-USER=US10723552 @CGN 1 1 734 @runat\_14022006\_125139\_12833 -NCPU=6 -ICPU=3  
-NO MAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WAE TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Genesep 21:\*  
1: Genesep1980s:\*  
2: Genesep1990s:\*  
3: Genesep2000s:\*  
4: Genesep2001s:\*  
5: Genesep2002s:\*  
6: Genesep2003as:\*  
7: Genesep2003bs:\*  
8: Genesep2004s:\*  
9: Genesep2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6099.5	41.7	1194	AAW39272	AAW39272 Porcine r
2	6038	41.3	1145	AAW32097	AAW32097 Miniature
3	6038	41.3	1145	AAW373286	AAW373286 Retroviral
4	6038	41.3	1145	AAW373447	AAW373447 Swine ret
5	5338	36.5	1737	AAW10044	AAW10044 MMLV gag-
6	5237	35.8	1784	AAW94427	AAW94427 FeLV F6A
7	5229	35.7	1784	AAW05898	AAW05898 Gene prod
8	4889	33.4	1069	AAW373284	AAW373284 Defective
9	4332.5	29.6	1203	AAW81572	AAW81572 Mus dunni

10	3981	27.2	1196	2	AAW75189	AAW75189 Osteoindu
11	3980.5	27.2	1199	7	ADH76470	ADH76470 Murine le
12	3979.5	27.2	1197	4	AAW49468	AAW49468 Canine re
13	3969.5	27.1	1199	3	AAW12994	AAW12994 MLV rever
14	3953	27.0	1224	2	AAW17947	AAW17947 MMLV pol
15	3850	26.2	744	8	ADS73450	ADS73450 Swine ret
16	3600	24.6	1079	2	AAW03163	AAW03163 MMLV reve
17	3534	24.1	666	8	ADS73449	ADS73449 Swine ret
18	3430	23.4	638	2	AAW32098	AAW32098 Miniature
19	3430	23.4	638	4	AAW70633	AAW70633 Porcine e
20	3430	23.4	638	4	AAW73287	AAW73287 Retroviral
21	3430	23.4	638	8	ADS73448	ADS73448 Swine ret
22	3425	23.4	638	4	AAW35114	AAW35114 PERV-C en
23	3412	23.3	638	9	ADY28028	ADY28028 Porcine e
24	3412	23.3	638	9	ADY28030	ADY28030 Porcine e
25	3412	23.3	638	9	ADY28036	ADY28036 Porcine e
26	3412	23.3	638	9	ADY28041	ADY28041 Porcine e
27	3412	23.3	638	9	ADY28038	ADY28038 Porcine e
28	3407	23.3	638	4	AAW70634	AAW70634 Porcine e
29	3318.5	22.7	678	9	ADY28032	ADY28032 Porcine e
30	3318.5	22.7	678	9	ADY28034	ADY28034 Porcine e
31	3214	22.0	1193	5	ABW79876	ABW79876 Spleen ne
32	3147.5	21.5	653	6	ABW82635	ABW82635 PERV enve
33	3135.5	21.4	653	6	ABW82636	ABW82636 PERV enve
34	2940.5	20.1	659	4	AAW35112	AAW35112 PERV-1-15
35	2879.5	19.7	661	9	ADY28039	ADY28039 Porcine e
36	2874.5	19.6	660	9	ADY28012	ADY28012 Porcine e
37	2871.5	19.6	660	9	ADY28004	ADY28004 Porcine e
38	2866.5	19.6	660	2	AAW85453	AAW85453 Pig endog
39	2866.5	19.6	660	4	AAW35113	AAW35113 PERV-A en
40	2866.5	19.6	660	4	AAW70632	AAW70632 Porcine e
41	2851	19.5	678	9	ADY28008	ADY28008 Porcine e
42	2849.5	19.5	660	9	ADY28014	ADY28014 Porcine e
43	2849.5	19.5	660	9	ADY28006	ADY28006 Porcine e
44	2835	19.4	678	9	ADY28010	ADY28010 Porcine e
45	2827	19.3	540	8	ADS73451	ADS73451 Swine ret

ALIGNMENTS

RESULT 1  
AAW39272  
ID AAW39272 standard; protein; 1194 AA.  
XX  
AC AAW39272;  
XX  
AC AAW39272;  
XX  
DT 27-AUG-2003 (revised)  
DT 19-MAY-1998 (first entry)  
XX  
DE Porcine retrovirus POL protein.  
XX  
KW Porcine retrovirus; PoEV; POL protein; ENV protein; GAG protein; vaccine;  
KW diagnosis; xenotransplantation; prophylactic; therapeutic.  
XX  
OS Pig endogenous retrovirus.  
XX  
PN WO9740167-A1.  
XX  
PD 30-OCT-1997.  
XX  
PF 18-APR-1997; 97WO-GB001087.  
XX  
PR 19-APR-1996; 96GB-00008164.  
PR 10-FEB-1997; 97GB-00002668.  
XX  
(QONE-) Q-ONE BIOTECH LTD.  
(IMUT-) IMUTRAN LTD.  
XX  
Galbraith DN, Haworth C, Lees GM, Smith KT;  
DR WPI; 1997-535851/49.  
DR N-PSDB; AAW09700.  
XX

PT Polynucleotide encoding porcine retrovirus expression product - useful to  
develop products for use in vaccines, diagnosis and xeno-transplantation.

PS Claim 6; Fig 3; 69pp; English.

XX This sequence represents the porcine retrovirus (PoRV) polymerase (POL)  
protein. This protein and other porcine retroviral proteins e.g. the  
CC virion core (GAG) and envelope (ENV) proteins can be used to develop  
CC viral vaccines, antisense nucleic acids, ribozymes and other antiviral  
CC agents. They can also be used in xeno-transplantation technology and as  
CC diagnostic tools. (Updated on 27-AUG-2003 to correct OS field.)

XX Sequence 1194 AA;

SQ Alignment Scores:

Pred. No.: 0 Length: 1194  
Score: 6099.50 Matches: 1154  
Percent Similarity: 98.5% Conservative: 19  
Best Local Similarity: 96.9% Mismatches: 17  
Query Match: 41.7% Indels: 1  
DB: 2 Gaps: 1

US-10-723-552-3 (1-8132) x AAW39272 (1-1194)

QY	2160	GGGAGACGGGTTTCGGACCCCTCCCGAGCCCGAGGTAACCTTTGAAGGTGGAGGGCAA	2219
DB	1	GIYARGGlySerAspProLeuProGluProArgValThrLeuLysValGluGlyGln	20
QY	2220	CCAGTTGAGTTCCTGGTTGATACCGGAGCGAAACATTCAGTGTACTACAGCCATTAGGA	2279
DB	21	ProValGluPheLeuValAspThrGlyAlaGluHisSerValLeuLeuGlnProLeuGly	40
QY	2280	AACTAAAGATAAAACTCTGGGTGATGGTGCACACGGGCAACACAGATTCATGG	2339
DB	41	LysLeuLysGluLysSerTrpValMetGlyAlaThrGlyGlnArgGlnTrpProTrp	60
QY	2340	ACTACCGAAGAACAGTTGAGTGGAGTGGAGCGGTAAACCCACTCGTTTCTGGTCATA	2399
DB	61	ThrThrArgargThrValAspLeuGlyValGlyArgValThrHisSerPheLeuValIle	80
QY	2400	CCTGAGTGCCAGCACCCCTCTTAGGTAGACTTATTGACCAAGATGGAGCACAAAT	2459
DB	81	ProGluCysProValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle	100
QY	2460	TCCTTTGAACAAGGAAACAGAACTGTCTGCAAAATAACAACTCATCTACTGTGTACC	2519
DB	101	SerPheGluGlnGlyArgProGluValSerValAsnAsnLysProIleThrValLeuThr	120
QY	2520	CTCCAAATTAGATACGAATATCGACTATCTCTCCCTAGTAAAGCTGATCAAAATATA	2579
DB	121	LeuGlnLeuAspAspGluTrpArgLeuTrpSerProGlnValLysProAspGlnAspIle	140
QY	2580	CAATTCCTGGTTGGAACAGTTTCCCAAGCCTGGGCGAGAAACCGCAGGATGGGTTGGCA	2639
DB	141	GlnSerTrpLeuGluGlnPheProGlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAla	160
QY	2640	AAGCAAGTCCCCCAGCTATTATCACTGAAGCGCAGTGCACACAGTGTCACTCAGTACA	2699
DB	161	LysGlnValProProGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArg	180
QY	2700	CAGTACCCCTTGATGAAGAGCTCAAGAGGAATTCGGCCGCATGTCCAAAGATTAATC	2759
DB	181	GlnTrpProLeuSerArgGluAlaArgGluGlyIleTrpProHisValGlnArgLeuIle	200
QY	2760	CAACAGGGCATCTAGTTCCTGCTCCCAATCTCCCTGGAAATCTCCCTGTCTACCGGTTAGA	2819
DB	201	GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLeuProValArg	220
QY	2820	AAGCTGGGACTAATGACTATCGACGATCAGAGCTTGAGAGAGGTCAATAACGGGTG	2879
DB	221	LysProGlyThrAsnAspTrpArgProValGlnAspLeuArgGluValAsnLysArgVal	240
QY	2880	CAGGATATACACCAACAGTCCCGAACCCCTTATAACCTCTGTGTCTCTCCCAACCCCAA	2939

DB	241	GlnAspIleHisProThrValProAsnProTrpAsnLeuLeuSerAlaLeuProProGlu	260
QY	2940	CGGAGCTGTATACAGATTGGACTTAAAGGATGCTTCTTCTGCTGAGATTACACCCC	2999
DB	261	ArgAsnTrpTrpThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisPro	280
QY	3000	ACTAGCAACCACTTTTTCGCTTCCGATCGAGATCCAGGTACGGGAAGAACCGGCGCAG	3059
DB	281	ThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln	300
QY	3060	CTCACCTGACCCGACTGCCCAAGGGTTCAAGAACTCCCGACCATCTTTGACGAAGCC	3119
DB	301	LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePheAspGluAla	320
QY	3120	CTACACAGAGACTGGCCAACTTCAGGATCCAAACACCCCTCAGTGACCTCTCCAGTAC	3179
DB	321	LeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnTrp	340
QY	3180	GTGGATGACTGCTCTGCGGGGAGCCACCAACAGAGACTGTCTAGAAGCGCAGGAAGCA	3239
DB	341	ValAspAspLeuLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGluGlyThrLysAla	360
QY	3240	CTACTGCTGGAATTTGTCTGACCTAGGCTACAGAGCCTCTGCTAAGAAAGCCCGATTTC	3299
DB	361	LeuLeuLeuGluLeuSerAspLeuGlyTyArgAlaSerAlaLysLysAlaGlnIleCys	380
QY	3300	AGAGAGAGGTAACTACTTGGGGTACAGTTTCGGGAGCGGCGAGGATGGCTGACGGAG	3359
DB	381	ArgArgGluValThrTrpLeuGlyTrpSerLeuArgGlyGlyGlnArgTrpLeuThrGlu	400
QY	3360	GCACGGAAGAAATCTAGTCCAGATACCGGCCCAACACACACCAACAAATGAGAGAG	3419
DB	401	AlaArgLysLysThrValGlnIleProAlaProThrThrAlaLysGlnValArgGlu	420
QY	3420	TTTTTGGGACAGCTGGATTTTCGACTGTGGATCCCGGGTTTCGCGACCTTAGCAGCC	3479
DB	421	PheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAlaAla	440
QY	3480	CACTCTACCGCTAAACAAAGAAAGGGGAATTCCTCTGGGCTCTGAGCACCAAGAG	3539
DB	441	ProLeuTrpProLeuThrLysGluLysGlyPheSerTrpAlaProGluHisGlnLys	460
QY	3540	GCATTTGATGCTATCAAAAGCCCTGCTGAGCGACCTGCTCTGCGCCCTCCCTGACGTA	3599
DB	461	AlaPheAspAlaLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspVal	480
QY	3600	ACTAAACCCCTTTTACCCCTTTATGTGGATGAGCGTAAAGGAGTAGCCCGGGAGTTTAAAC	3659
DB	481	ThrLysProPheThrLeuTrpValAspGluArgLysGlyValAlaArgGlyValLeuThr	500
QY	3660	CAAAACCTTAGGACCATGGAGAGACCTGCTGCGCTACTCTGCTCAAGAAAGCTCGATCT	3719
DB	501	GlnThrLeuGlyProTrpArgProValAlaTyLeuSerLysLysLeuAspProVal	520
QY	3720	GCAGTGGTTGGCCCATATGCTGAGGCTATCGCAGCTGTGGCCATCTACTGTGCAAGGAC	3779
DB	521	AlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaValAlaLeuValLysAsp	540
QY	3780	GCTGCAAAATTTGACTTTGGGACAGATAATACTGTAAATAGCCCCCTCATGATTGGAGAAC	3839
DB	541	AlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProHisAlaLeuGluAsn	560
QY	3840	ATCGTTTCGGAGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCATCTCAAGC	3899
DB	561	IleValArgGlnProProAspArgTrpMetThrAsnAlaArgMetThrHisTrpGlnSer	580
QY	3900	CTGCTTCTCACAGAGAGGTTCAGTTTCGCTCCACACGCGCTCTCAACCTGCGACTCTT	3959
DB	581	LeuLeuLeuThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu	600
QY	3960	CTGCTTGAAGAGACTGATGAACCAAGTCACTGATTGCGCATCAACTATTGATTGAGGAG	4019

Db	601	LeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLeuLeuGluGlu	620
QY	4020	ACTGGGTCGCAAGACCTTACAGACATACCCTGCTGAGAGAGTGTAACTCTGGTTC	4079
Db	621	ThrGlyValArgLysAspLeuThrAspIleProLeuThrGlyGluValLeuThrTrpPhe	640
QY	4080	ACTGACGGAGCAGCTATGTGGTGGAGGTAAAGAGATGGCTGGCGGGTGGTGGAC	4139
Db	641	ThrAspGlySerSerTyrValValGluGlyLysArgMetAlaGlyAlaValValAsp	660
QY	4140	GGGACCCGACACATCTGGGGCAGCAGCTGCCGGAAGGAACCTTCAGCACAAAAGCTGAG	4199
Db	661	GlyThrArgThrIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu	680
QY	4200	CTCATGGCCCTCAGCAAGCTTTGCGGCTGGCGAAGGAAATCCATAACATTTATACG	4259
Db	681	LeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlySerIleAsnIleTyrThr	700
QY	4260	GACAGCAGGTATGCCCTTTGGGACTGCACAGTCATATGGGGCCATCTATAACAAAGGGGG	4319
Db	701	AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrLysGlnArgGly	720
QY	4320	TTGCTTACCTCAGCAGGGAGGAATAAAGAACAAAGAGGMAATCTTAAGCCCTATTAGAA	4379
Db	721	LeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluIleLeuSerLeuLeuGlu	740
QY	4380	GCGGTACATTTACCAAAAGGCTAGCTATTATACACTGCTCTGGACATCAGAAAGCTAAA	4439
Db	741	AlaLeuHisLeuProLysArgLeuAlaIleIleHisCysProGlyHisGlnLysAlaLys	760
QY	4440	GATCTCATATCCAGAGAAACAGATGGCTGACCCGGTGGCCAAAGCAGGCCAGGCT	4499
Db	761	AspLeuIleSerArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaGlnAla	780
QY	4500	GTTAACTTCTGCTATATAAGAAATGCCCAAGCCCCAGAACCCAGACGACAGTACACC	4559
Db	781	ValAsnLeuLeuProIleLeuThrProLysAlaProGluProArgGlnTyrThr	800
QY	4560	CTGAGACTGCGCAGAGATAAAAAGATAGACCACTTCTCAGACTCCGGAAGGGACC	4619
Db	801	LeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGlyThr	820
QY	4620	TGCTATACCTCAGATGGGAAGAAATCTTGCCGCCAACAAAGAGGGTTAGAAATGTCCAA	4679
Db	821	CysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeuGluTyrValGln	840
QY	4680	CAGATACATCGTCTAACCCACTAGCACTTAACACTGCAGCAGTGTGTCAGACATCC	4739
Db	841	GlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGlnLeuValArgThrSer	860
QY	4740	CCTTATCATGTTCTGAGGCTACAGGAGTGGCTGACTCGGTGGTCAAACTTGTGTGCC	4799
Db	861	ProTyrHisValLeuArgLeuProGlyValAlaAspSerValValLysHisCysValPro	880
QY	4800	TGCCAGCTGGTAAATCTTAATCCTTCCAGAATGCCCTCCAGGAAGAGACTAAGGGGAAGC	4859
Db	881	CysGlnLeuValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySer	900
QY	4860	CACCAGGCGCTCAGTGGGAAGTGCATTCACCTGAGGTAAAGCCGGCTAAATACCGAAC	4919
Db	901	HisProGlyAlaHisIleTrpGluValAspPheThrGluValLysProAlaLysTyrGlyAsn	920
QY	4920	AAATACCTATTGCTTTTCTAGACACCTTTTTCAGGATGGGTAGAGCTTATCTCTACTAAG	4979
Db	921	LysTyrLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaTyrProThrLys	940
QY	4980	AAAGAGACTTCAACCGTGGTGGCTAAAAAATACTCGAAGAAATTTTCCAGATTTGGA	5039
Db	941	LysGluThrSerThrValValAlaLysLysIleLeuGluGluIlePheProArgPheGly	960
QY	5040	ATACCTAAGGTATAGGTACAGCAATGCTCAGCTTTTGTGCCCAGGTAAAGTCAGGA	5099
Db	961	IleProLysValIleGlySerAspAsnGlyProAlaPheValAlaGlnValSerGlnGly	980
QY	5100	CTGGCCAGATATTGGGATTTGATTGGAAACTGCATTGTGCATACAGACCCCAAAGCTCA	5159
Db	981	LeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSer	1000
QY	5160	GGACAGGTAGAGAGATGAATAGAACCATTAAGAGACCCCTTAACTAAATGACCGGGAG	5219
Db	1001	GlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThrThrGlu	1020
QY	5220	ACTGGGTTAAATGATTGGATAGCTCTCTCCCTCCCTTTGTCTTTTATAGGTTAGGACACC	5279
Db	1021	ThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeuPheArgValArgAsnThr	1040
QY	5280	CCTGGACACTTTGGGCTGACCCCTTATGAATTACTCTACGGGGGACCCCTTCATTGGTA	5339
Db	1041	ProGlyGlnPheGlyLeuThrProTyrGluLeuLeuTyrGlyGlyProProProLeuAla	1060
QY	5340	GAATTTGCTTCTGTACATAGTCTGACGTGCTGCTTTTCCAGCCTTTGTTCTTAGGCTC	5399
Db	1061	GluIleAlaPheAlaHisSerAlaAspValLeuLeuSerGlnProLeuPheSerArgLeu	1080
QY	5400	RAGGCACCTTGAGTGGGTGAGACAACGAGCGTGGAGCACTCCGGGAGGCTACTACGGA	5459
Db	1081	LysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTyrSer---	1099
QY	5460	GGAGGAGACTTCAGATCCACATCGTTTCCAAGTGGGAGATTTCAGTCTACGTTAGACGC	5519
Db	1100	GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTyrValArgArg	1119
QY	5520	CACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGGCCCTTATCTCGTACTTTTGACCACA	5579
Db	1120	HisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProTyrLeuValLeuLeuThrThr	1139
QY	5580	CCACGGCTGTCAAAGTCGAAGGAATCTCCACCTGCATCCATGCATCCACCGTTAAACCG	5639
Db	1140	ProThrAlaValLysValGluGlyIleProThrTrpIleHisAlaSerHisValLysPro	1159
QY	5640	GGCCACCTCCCGATTTCGGGGTGGAAAGCCGAAAGACTCGAAAATCCCTTAAAGCTTCGC	5699
Db	1160	AlaProProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeuLysLeuArg	1179
QY	5700	CTCCATCGGTGGTTCCTTACTCTCTCAATAAC	5732
Db	1180	LeuHisArgLeuValProTyrSerAsnAsnAsn	1190
RESULT 2			
AAW32097			
ID	AAW32097	standard; protein; 1145 AA.	
AC	AAW32097;		
XX			
DT	27-AUG-2003	(revised)	
DT	09-FEB-1998	(first entry)	
XX			
DE	Miniature swine retrovirus POL protein.		
XX			
KW	Retrovirus; porcine; POL protein; xenotransplantation; infectious;		
KW	provirus; organ transplant; donor; activated virus; PCR.		
OS	Pig endogenous retrovirus.		
XX			
FH	Key	Location/Qualifiers	
FT	Protein	1..1145	
FT		/label= POL_protein	
XX			
PN	WO9721836-A1.		
XX			
PD	19-JUN-1997.		
XX			
PF	13-DEC-1996;	96WO-US019680.	
XX			
PR	14-DEC-1995;	95US-00572645.	
XX			

PA (GEHO ) GEN HOSPITAL CORP.  
 XX Fishman JA;  
 PI WPI; 1997-332804/30.  
 XX DR N-PSDB; AAT74884.  
 XX  
 XX New nucleic acid from porcine retroviruses - used for detecting viruses  
 PT in transplant or other tissue and for assessing risk of transmitting  
 PT infection to graft recipient.  
 XX  
 XX Claim 22; Fig 3; 128pp; English.  
 XX  
 CC This is a porcine retrovirus from miniature swine containing the coding  
 CC region for a putative viral POL protein. This sequence and PCR fragments  
 CC generated from the sequence (see AAF74812-174882) could be used to screen  
 CC organs for porcine retroviruses prior to xenotransplantation.  
 CC Transplantation can increase the likelihood of retroviral activation if  
 CC intact and infectious proviruses are present. The porcine retroviral  
 CC sequence can be used to generate probes to determine the level (e.g. copy  
 CC number) of intact (i.e. potentially replicating) porcine provirus  
 CC sequences in a strain of xenograft transplantation donors. It can be used  
 CC to detect mutations, genetic lesions or viral recombinants and also to  
 CC determine the histological localisation of activated retrovirus. Using  
 CC Polymerase Chain Reaction DNA Quantitation (PDQ) on blood mononuclear  
 CC cells, infectivity titration and susceptibility testing can be performed.  
 CC Ultimately animal donors without intact porcine retroviral sequences or a  
 CC lower copy number of viral elements could be selected. (Updated on 27-AUG  
 CC -2003 to correct OS field.)  
 XX  
 XX Sequence 1145 AA;  
 SQ

Alignment Scores:  
 Pred. No.: 0 Length: 1145  
 Score: 6038.00 Matches: 1145  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 41.3% Indels: 0  
 DB: 2 Gaps: 0

US-10-723-552-3 (1-8132) x ANW32097 (1-1145)

QY	2307	ATGGGTGCCACAGGCAACAAACAGTATCCATGGACTACCGAAGAACAGTTGACTTGGGA	2366
DB	1	MetGlyAlaThrGlyGlnGlnGlnTyrProTyrThrThrArgargThrValaspLeuGly	20
QY	2367	GTGGGACGGGTAAACCCACCTGTTTCTGGTCATACCTGAGTGGCCAGACCCCTCTTAGGT	2426
DB	21	ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly	40
QY	2427	AGAGACTTATTGACCAAGATGGGAGCACAAATTTCTTTTGMACAGGGAAACAGAGTG	2486
DB	41	ArgaspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyLysProGluVal	60
QY	2487	TCTGCAATTAACAAACCTATCACTGTGTTCACCCCTCCAAATTAGATGACCAATATCGACTA	2546
DB	61	SerAlaasnAsnLysProIleThrValLeuThrLeuGlnLeuaspaspGluThrArgLeu	80
QY	2547	TACTCTCCCTAGTAAAGCCTGATCAAAATATACAAATTCGTGGTGAACAGTTTCCCCAA	2606
DB	81	TyrSerProLeuValLysProaspGlnAsnIleGlnPheThrLeuGluGlnPheProGln	100
QY	2607	GCTTGGGCGAAGAACCCGAGGATGGTTGGCAAGCAAGTTCCTCCCAACAGTTATTCAA	2666
DB	101	AlaThrPalaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln	120
QY	2667	CTCAAGCGCAGTCCACACACAGTGTCAAGTACAGAGTACCCCTTGAGTAAAGAGCTCAA	2726
DB	121	LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerLysGluAlaGln	140
QY	2727	GAAGGAATTCGGCCGATGTCCTCAAGATTAATCCAAACAGGGCATCTAGTCTCTGCCAA	2786
DB	141	GluglyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln	160

QY	2787	TCTCCCTGGAATACTCCCTGCTACCGGTTAGAAAGCCTGGGACTAATGACTATCGACCA	2846
DB	161	SerProTyrAsnThrProLeuLeuProValAlaGlyLysProGlyThrAsnAspTyrArgPro	180
QY	2847	GTACAGGACTTGAGAGAGGTCAATAAACGGGTGCAGGATATACACCAACAGTCCCGAAC	2906
DB	181	ValGlnaspLeuArgGluValasnLysArgValGlnAspIleHisProThrValProasn	200
QY	2907	CCTTATAACCTCTGTGTGCTCTCCACCCCAACCGAGCTGGTATACAGTATTGGACTTA	2966
DB	201	ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTyrThrValLeuAspLeu	220
QY	2967	AAGGATGCTCTTCTTCTGCTGAGATTACACCCCACTAGCAACCACTTTTTCCTTCGAA	3026
DB	221	LysaspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu	240
QY	3027	TGAGAGATTCACAGGTACCGGAAACACCGGCAGCTCACCTGGACCCGACTGCCCCAAAGG	3086
DB	241	TyrArgaspProGlyThrGlyArgThrGlyGlnLeuThrTyrThrArgLeuProGlnGly	260
QY	3087	TTCAAGAACTCCCGACCATCTTTTACGAAGCCCTACACAGAGACTGGCCAACTTCAGG	3146
DB	261	PheLysAsnSerProThrIlePheaspGluAlaLeuHisArgaspLeuAlaAsnPheArg	280
QY	3147	ATCCAAACCCCTCAGGTGACCTCTCCAGTAGCTGAGTACCTGCTCTGGCGGGAGCC	3206
DB	281	IleGlnHisProGlnValThrLeuLeuGlnTyrValaspLeuLeuLeuAlaGlyAla	300
QY	3207	ACCAACAGGACTGCTTAGAAGCACGAAGGCACACTACTGCTGGAATGCTGACCTAGGC	3266
DB	301	ThrLysGlnaspCysLeuGluGlyThrLysAlaLeuLeuLeuLeuSeraspLeuGly	320
QY	3267	TACAGAGCTCTGCTTAAGAAGCCCAAGATTTCAGGAGAGAGTAAACATCTTGGGGTAC	3326
DB	321	TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr	340
QY	3327	AGTTTTCGGGACGGCGAGGTGGCTGACGGAGGCACGGAAGAAACTGTAGTCCAGATA	3386
DB	341	SerLeuArgaspGlyGlnArgTyrLeuThrGluAlaargLysLysThrValValGlnIle	360
QY	3387	CGGGCCCCAACACACAGCCAAACAAATGAGAGAGTTTTTGGGACACAGCTGGATTTTCAGA	3446
DB	361	ProAlaProThrThrAlaLysGlnMetArgGluPheLeuGlyThrAlaGlyPheCysArg	380
QY	3447	CTGTGATCCCGGGTTTCGACCTTAGCAGCCCACTTACCCGCTAACCAAGAAAAA	3506
DB	381	LeuTyrIleProGlyPheAlaThrLeuAlaProLeuTyrProLeuThrLysGluLys	400
QY	3507	GGGGAATTCCTCTGGGCTCTGAGCACCAAGAGGCATTTGATGCTATCAAAAAGGCCCTG	3566
DB	401	GlyGluPheSerTyrPalaProGluHisGlnLysAlaPheaspAlaIleLysLysAlaLeu	420
QY	3567	CTGAGCGCACCTGCTCTGCGCCCTCCCTGACGTAACTAAACCCCTTTATGTGGAT	3626
DB	421	LeuSerAlaProAlaLeuAlaLeuProaspValThrLysProPheThrLeuTyrValasp	440
QY	3627	GAGCGTAAAGGAGTAGCCCGGGAGTTTAAACCAAAACCTTAGGACCATGGAGAACCT	3686
DB	441	GluArgLysGlyValAlaargGlyValLeuThrGlnThrLeuGlyProTyrArgArgPro	460
QY	3687	GTCGCTTACTCTGCAAGAGCTCCGATCTCTGTAGCCAGTGGTGGCCCATATGCTCTGAAG	3746
DB	461	ValAlaTyrLeuSerLysLysLeuaspProValAlaSerGlyTyrProIleCysLeuLys	480
QY	3747	GCTATCGCAGCTGTGGCCATCTGCTCAAGGACGCTGACAAATTTGACTTTGGACAGAAT	3806
DB	481	AlaIleAlaAlaValaIleLeuValLysaspAlaaspLysLeuThrLeuGlyGlnAsn	500
QY	3807	ATAACTGTATAGCCCCCATGATTGGAGAACATGCTGTCGGCAGCCCCCAGACCCGATGG	3866
DB	501	IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProaspArgTyr	520

QY 3867 ATGACCAACGCCCGCATGACCCACTATCAAGCCTGCTTCTCAGAGAGGGTCACGTTCC 3926  
DB 521 MetThrAsnAlaargMetThrHisTyrglnSerLeuLeuLeuThrGluArgValThrPhe 540  
QY 3927 GCTCCACCGACCGCTCTCAACCCCTGCCACTCTTCTGCTCAAGAGACTGATGAACAGTG 3986  
DB 541 AlaProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 560  
QY 3987 ACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGAC 4046  
DB 561 ThrHisAspCysHisGlnLeuLeuLeuGluGluThrGlyValarglysAspLeuThrAsp 580  
QY 4047 ATACCCGCTGACTGGAGAGTCTAACTGTTCACTGACGGAAGCAGCTATGTGTGAA 4106  
DB 581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSertyrValValGlu 600  
QY 4107 GGTAAAGAGGATGCTGGGGCGGCTGTGAGCGGACCGCACGATCTGGGCCACGACG 4166  
DB 601 GlylysArgMetAlaGlyAlaIaValIaAspGlyThrargThrIleTrpAlaSerSer 620  
QY 4167 CTGCCGGAAGGAATCTCAGCACAAAAGGCTGAGCTCATGGCCCTCAGCAAGCTTTGCCG 4226  
DB 621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
QY 4227 CTGCCGGAAGGAATCCATAACATTTATACGACAGCAGGTATGCTTTGGCACTGCA 4286  
DB 641 LeuAlaGluGlyLysSerIleAsnIleTyThrAspSerArgTyThrAlaPheAlaThrAla 660  
QY 4287 CAGTCATGGGGCCCATCTATAACAAAGGGGTGCTTACCTCAGCAGGAGGGGAATA 4346  
DB 661 HisValHisGlyAlaIleTyLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
QY 4347 AAGAACCAAGAGGAAATCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCT 4406  
DB 681 LysAsnLysGluGluLeuSerLeuLeuGluAlaValHisLeuProLysArgLeuAla 700  
QY 4407 ATTATACATGCTCTGGACATCAGAAAGCTAAGATCTCATATCCAGAGAAACAGATG 4466  
DB 701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720  
QY 4467 GCTGACGGGTGCCAAGCAGCGAGCCAGGGTGTAACTTCTGCTTATATAGAAATG 4526  
DB 721 AlaAspArgValAlaLysGlnAlaLagInGlyValAsnLeuProIleIleGluMet 740  
QY 4527 CCCAAGCCCGCAAGACCCAGACAGACTACACCTAGAAGACTGGCAAGAGATAAAAAAG 4586  
DB 741 ProLysAlaProGluProArgGlnTyThrLeuGluAspTrpGlnGluIleLysLys 760  
QY 4587 ATAGACCACTTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGGAATC 4646  
DB 761 IleAspGlnPheSerGluThrProGluGlyThrCysTyThrSerAspGlyLysGluIle 780  
QY 4647 CTGCCCCACAAAGAGGGTTAGATATGTCACAGATACATCGTCTAACCCACCTAGGA 4706  
DB 781 LeuProHisLysGluGlyLeuGluTyValGlnGlnIleHisArgLeuThrHisLeuGly 800  
QY 4707 ACTAAACACTGCAGCAGTTGTCAGAACATCCCTTATCATGTTCTGAGGCTACCAGGA 4766  
DB 801 ThrLysHisLeuGlnGlnLeuValargThrSerProTyThrHisValLeuargLeuProGly 820  
QY 4767 GTGGCTGACTCGGTGTCAAAATGTTGTGTCCTGCCAGCTGTTAATGCTTAATCTTCC 4826  
DB 821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
QY 4827 AGAATGCTCCAGGAGAGACTAAGGGAAGCCACCGGCGCTCACTGGGAAGTGGAC 4886  
DB 841 ArgMetProProGlyLysArgLeuargGlySerHisProGlyAlaHisTrpGluValAsp 860  
QY 4887 TTCACTGAGTAAAGCCGCTAAATACGGAACAAATACTATTGTTTCTGTAGACACC 4946  
DB 861 PheThrGluValLysProAlaLysTyThrGlyAsnLysTyThrLeuLeuValPheValAspThr 880  
QY 4947 TTTTCAGGATGGGTAGAGCTTATCTCTACTAAGAAAGAGACTTCAACCGTGTGGCTAAA 5006

DB 881 PheSerGlyTrpValGluAlaTyProThrLysLysGluThrSerThrValValAlaLys 900  
QY 5007 AAAATACTGGAGAAATTTTCCAAAGATTGGTAATACCTTAAGCTAATAGGGTCAGACAAT 5066  
DB 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
QY 5067 GTTCCAGCTTTTGTCTCCCAAGTAAGTCAGGGACTGCGCAAGATATTGGGGATTGATTGG 5126  
DB 921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
QY 5127 AAATCTGATTGTGCTATACAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAAC 5186  
DB 941 LysLeuHisCysAlaTyArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
QY 5187 ATTAAAGAGACCTTACTAAATTGACCGCGGAGACTGGCGTTAATCATTTGGATAGCTCTC 5246  
DB 961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTrpIleAlaLeu 980  
QY 5247 CTGCCCTTTGTCTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCTAT 5306  
DB 981 LeuProPheValLeuPheArgValAlaArgAsnThrProGlyGlnPheGlyLeuThrProTy 1000  
QY 5307 GAATTAATCTACGGGGACCCCTCCATTTGGTAGAAATTTGCTTCTCATAGTGTCTGAC 5366  
DB 1001 GluLeuLeuTyrglyGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1020  
QY 5367 GTGCTGCTTTCCCGCCTTTGTTCTTAGGCTCAGGCACTTGGTGGTGAGACAACGA 5426  
DB 1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
QY 5427 GCGTGAGGACAACTCCGGGAGCGCTACTCAGGAGGAGAGACTTGAGATCCACATCGT 5486  
DB 1041 AlaTrpArgGlnLeuArgGluAlaTySerGlyGlyAspLeuGlnIleProHisArg 1060  
QY 5487 TTCCAAGTGGGAGATTCTAGTTAGACCCCGCTGAGGAAACCTTCGAGACTCGG 5546  
DB 1061 PheGlnValGlyAspSerValTyValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
QY 5547 TCGAAGGCGCTTATCTCGTACTTTTACCACACCGCTGTGAAAGTCGAGAAATC 5606  
DB 1081 TrpLysGlyProTyLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1100  
QY 5607 TCCACTGGATCCATCCATCCACAGTTAAACCGGCGCACCTCCCGATTCCGGGTGAAA 5666  
DB 1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProProAspSerGlyTrpLys 1120  
QY 5667 GCCGAAAAGACTGAAAATCCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTC 5726  
DB 1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTySerVal 1140  
QY 5727 AATAACCTCTCAGAC 5741  
DB 1141 AsnAsnLeuSerAsp 1145

## RESULT 3

AAB73286  
ID AAB73286 standard; protein; 1145 AA.

XX AAB73286;

XX 23-MAY-2001 (first entry)

DE Retroviral protein #2 found in miniature swine.

XX Retrovirus; graft tranplantation; xenotransplantation; miniature swine.

OS Unidentified.

XX US6190861-B1.

XX 20-FEB-2001.

XX



PF 13-DEC-1996; 96US-00766528.  
 XX  
 PR 14-DEC-1995; 95US-00572645.  
 XX  
 PA (GEOH) GEN HOSPITAL CORP.  
 XX Fishman JA;  
 XX  
 PI  
 XX  
 XX WPI; 2001-256211/26.  
 DR N-PSDB; AAF77727.  
 XX  
 XX  
 PT Assessing risk of endogenous retroviruses in clinical practice and in  
 PT xenotransplantation, comprises using probe sequences derived from swine  
 PT or miniature swine retroviral genome.  
 XX  
 XX  
 PS Disclosure; Fig 3; 127pp; English.  
 XX  
 CC The present invention relates to a method for screening a cell or tissue  
 CC for the presence or expression of a retrovirus (RV), comprising  
 CC contacting a target nucleic acid from the cell or tissue with a second  
 CC nucleic acid from the present invention (e.g. AAF77727 or a fragment  
 CC thereof). The method is useful for RV detection and to assess graft  
 CC transplantation risk. Screening of animals allows the elimination of  
 CC donors with active replication of known viruses. Inactive proviruses can  
 CC be detected and inactivated, allowing identification and elimination of  
 CC potential human pathogens derived from swine in a manner not possible in  
 CC the outbred human organ donor population and is important to the  
 CC development of human xenotransplantation  
 XX  
 SQ Sequence 1145 AA;  
 XX

Alignment Scores:  
 Pred. No.: 0 Length: 1145  
 Score: 6038.00 Matches: 1145  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 41.3% Indels: 0  
 DB: 4 Gaps: 0

US-10-723-552-3 (1-8132) x AAB73286 (1-1145)

QY 2307 ATGGGTGCCACAGGCAACACAGTATCCATGGACTACCCGAAAGACAGTTGACTTGGGA 2366  
 Db 1 MetGlyAlaThrGlyGlnGlnTyrProTyrThrArgArgThrValAspLeuGly 20  
 QY 2367 GTGGGACGGTAACCCACTCGTTTCGTGTCATACCTGAGTGGCCAGCACCCCTCTTAGT 2426  
 Db 21 ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 40  
 QY 2427 AGAGACTTATTGACCAAGATGGGAGCACAAATTTCTTTGAACAGGAAACCCAGAGTG 2486  
 Db 41 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyProGluVal 60  
 QY 2487 TCTGCAATAACAAACCTATCACTGTGTGACCTCCAAATTAGATGACCAATATCGACTA 2546  
 Db 61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspGluTyrArgLeu 80  
 QY 2547 TACTCTCCCTAGTAAGCCTGATCAAAATATACAAATTCGTGTTGGAAACAGTTTCCCCAA 2606  
 Db 81 TyrSerProLeuValLysProAspGlnAsnIleGlnPheTrpLeuGluGlnPheProGln 100  
 QY 2607 GCCTGGGCAGAACCCGAGGTGGTTGGCAAGCCAGATTCCTCCCAAGTTATTCAA 2666  
 Db 101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProGlnValIleGln 120  
 QY 2667 CTGAAGCCAGTCCCAACACAGTGTGAGTCAGTCAGACAGTACCCCTTGAGTAAAGAGCTCAA 2726  
 Db 121 LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerLysGluAlaGln 140  
 QY 2727 GAAGGAATTGCGCGCATGTCCTCAAGATTAATCCACAGGGCATCTAGTCTCTGTCTCAA 2786  
 Db 141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 160

QY 2787 TCTCCCTGGAATACTCCCTGCTACCGGTTAGAAAGCCTGGGACTAATGACTATCGACCA 2846  
 Db 161 SerProTyrAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 180  
 QY 2847 GTACAGGACTTGAGAGAGGTCAATAAACCGGTGCAGGATATACACCAACAGTCCCGAAC 2906  
 Db 181 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200  
 QY 2907 CCTTATAACCTCTGTGTGCTCTCCACACCCCAACGGAGCTGGTATACAGTATTGGACTTA 2966  
 Db 201 ProTyrAsnLeuLeuCysAlaLeuProGlnArgSerTrpTyrThrValLeuAspLeu 220  
 QY 2967 AAGGATGCTCTTCTGCTGAGATTACACCCCACTAGCAACCACTTTTGTCTTCGAA 3026  
 Db 221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240  
 QY 3027 TGGAGAGATCCAGGTACGGGAACCCGGGAGCTCACTGGACCCGACTGCTCCCAAGGG 3086  
 Db 241 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260  
 QY 3087 TTCAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCACTTCAGG 3146  
 Db 261 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280  
 QY 3147 ATCCAAACCCCTCAGGTGACCCCTCTCCAGTACGTGGATGACCTGCTCTGGCGGAGCC 3206  
 Db 281 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla 300  
 QY 3207 ACCAAACAGGACTGTCTTAGAGGACCAAGGCACCTACTCTGGAATTGTCTGACTAGGC 3266  
 Db 301 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuLeuSerAspLeuGly 320  
 QY 3267 TACAGAGCTCTGTCTAAGAGGCCAGATTTCCAGGAGAGAGTACATCTTGGGGTAC 3326  
 Db 321 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 340  
 QY 3327 AGTTTCGGCGAGCGGCGAGCATGGCTGACGGAGGCCAGGAAGAAAACTGTAGTCCAGATA 3386  
 Db 341 SerLeuArgAspGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 360  
 QY 3387 CCGGCCCCAACCAAGCCAAACAAATGAGAGAGTTTTTGGGGACAGCTGGATTTTGAGA 3446  
 Db 361 ProAlaProThrThrAlaLysGlnMetArgGluPheLeuGlyThrAlaGlyPheCysArg 380  
 QY 3447 CTGTGGAATCCCGGGTTTCGGACCTTAGCAGCCCACTCTACCCGCTACCAACCAAGAAAA 3506  
 Db 381 LeuTrpIleProGlyPheAlaThrLeuAlaLeuProLeuTyrProLeuThrLysGluLys 400  
 QY 3507 GGGGAATTCCTCTGGGCTCTGAGCACCAAGAGCAATTTGATGCTATCAAAAAGGCCCTG 3566  
 Db 401 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420  
 QY 3567 CTGAGCGCACCTGTCTGCGCCCTCCTGACGTAACTAAACCCCTTTTACCCTTTATGTGGAT 3626  
 Db 421 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 440  
 QY 3627 GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTAGGACCATCGGAAGACCT 3686  
 Db 441 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 460  
 QY 3687 GTCGCTTACTGTCAAAGAAAGCTCGATCCTGTAGCCAGTGGTTGGGCCCATATGCTCTGAAG 3746  
 Db 461 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLys 480  
 QY 3747 GCTATCCGAGCTGTGGCCATCTGCTCAAGGACGTGACAAATGTGCTTTGGGACAGAAT 3806  
 Db 481 AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 500  
 QY 3807 ATAACTGTAAATAGCCCCCATGCTTGGAGACATCGTTCCGCGAGCCCCCAGACCCATGG 3866  
 Db 501 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 520  
 QY 3867 ATGACCAACGCGCGCATGATGATCAAGCCCTGCTTCTCTACAGAGAGGGTCACGTTTC 3926

Db 521 MetThrAsnAlaArgMetThrHisLeuGlnSerLeuLeuLeuThrGluArgValThrPhe 540  
QY 3927 GCTCCACCGCCCTCTCAACCCCTGCACCTCTTCCTGCTGAAGAGACTGATGAACAGATG 3986  
Db 541 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuLeuProGluGluThrAspGluProVal 560  
QY 3987 ACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGAC 4046  
Db 561 ThrHisAspCysHisGlnLeuLeuLeuLeuGluThrGlyValArgLysAspLeuThrAsp 580  
QY 4047 ATACCCCTGACTGGAGAGTCTAACTGGTTCACTGACGGAAGAGCTATGTGGTGA 4106  
Db 581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 600  
QY 4107 GGTAAAGAGATGCTCGGGCGGGTGGTGGAGCGGACCGCACGATCTGGGCCACGAC 4166  
Db 601 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 620  
QY 4167 CTGCCGGAAGGAATCTCAGCACAAAAGGCTGAGCTCATGGCCCTCAGCAAGCTTTGCGG 4226  
Db 621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
QY 4227 CTGCCGGAAGGAATCCATAACATTTATACGGAAGAGTATGCCCTTTCGACTGCA 4286  
Db 641 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 660  
QY 4287 CAGTACATGGGGCCATCTATAACAAAGGGGTGCTTACCTCAGCAGGGGGGAATA 4346  
Db 661 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
QY 4347 AAGAAACAAAGAGAAATCTAAGCCCTATTAGAAGCCGTATACATTTACAAAAGGCTAGCT 4406  
Db 681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaValHisLeuProLysArgLeuAla 700  
QY 4407 ATTATACATGCTCTGGACATCGAAAGCTAAAGATCTCATATTCGAGAGAAACCGATG 4466  
Db 701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720  
QY 4467 GCTGACCGGGTTCGCAAGCAGGCGCCAGGGTGTAACTTCTGCTATATAGAAATG 4526  
Db 721 AlaAspArgValAlaLysGlnAlaValGlnGlyValAsnLeuLeuProIleIleGluMet 740  
QY 4527 CCCAAAGCCCGAAGCCAGACGACAGTACACCTCAGAGAGACTGGCAAGAGATAAAAAG 4586  
Db 741 ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760  
QY 4587 ATAGACCACTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATC 4646  
Db 761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 780  
QY 4647 CTGCCCCCAAGAGAGGGTTAGAAATATGTCCAAGATACATCGTCAACCCACCTAGGA 4706  
Db 781 LeuProHisLysGlyGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
QY 4707 ACTAAACACCTGCAGCAGTTGCTCAGAACATCCCTTATCATGTTCTGAGGCTACAGGA 4766  
Db 801 ThrLysHisLeuGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
QY 4767 GTGGCTGACTCGGTGTTCAAAATGTTGTGCCCTCCAGCTGGTTAATGCTTAATCTTCC 4826  
Db 821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
QY 4827 AGAATGCTCCAGGAGAGACTAAGGGAAGCCACCCAGGCGCTCACTGGGAAGTGAC 4886  
Db 841 ArgMetProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860  
QY 4887 TTCCTGAGGTAAAGCCGCTAAATACGGAACAAATACCTATTGTTTCTAGACACC 4946  
Db 861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
QY 4947 TTTTCAGGATGGGTAGAGCTTATCTCTACTAAGAAAGAGACTTCAACCGCTGGCTAAA 5006

Db 881 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900  
QY 5007 AAAATACTCGAAGAAATTTTCCACGATTGGGAATACCTAAAGTAAATAGGGTCAGACAAT 5066  
Db 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
QY 5067 GGTCCAGCTTTTGTTCGCCAGGTAACTCAGGGACTCGGCAAGATATTGGGGATTGATTGG 5126  
Db 921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
QY 5127 AAATGTCATTGTCATACAGACCCCAAGCTCAGGACAGTAGAGAGGATGAATAGAAC 5186  
Db 941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
QY 5187 ATTAAGAGACACCTTACTAAATTTGACCGGAGACTGGCGTTAATCATGATTGGATAGCTCTC 5246  
Db 961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTrpIleAlaLeu 980  
QY 5247 CTGCCCTTTGTGCTTTTAGGGTTAGGAACACCCCTGGACAGCTTTCGGGCTGACCCCTAT 5306  
Db 981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
QY 5307 GAATTACTCTACGGGGACCCCTCCATTTGGTAGAAATTTGCTTCTCTACATAGTCTGAC 5366  
Db 1001 GluLeuLeuTyrGlyGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1020  
QY 5367 GTGCTGCTTTCCAGCCCTTTGCTCTAGGCTCAGGCACTTTCAGTGGGTGAGACACGA 5426  
Db 1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
QY 5427 GGTGGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGAGACTTGCAGATCCACATCGT 5486  
Db 1041 AlaTrpArgGlnLeuArgGluAlaTyrSerGlyGlyAspLeuGlnIleProHisArg 1060  
QY 5487 TTCCAAGTCGGAGATTACGTTCAGTTAGACGCCACCGTCGAGGAAACCTCGAGACTCGG 5546  
Db 1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
QY 5547 TCGAAGGGCCCTTATCTGCTACTTTTGACCAACCAACGGCTGTGAAGTCGAAGGAATC 5606  
Db 1081 TrpLysGlyProTyrLeuValLeuLeuThrProThrAlaValLysValGluGlyIle 1100  
QY 5607 TCCACTTGTGATCCATCCACCGTTAAACCGGCGCCACCTCCCGATTTCGGGGTGGAAA 5666  
Db 1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTrpLys 1120  
QY 5667 GCCAAAAGACTGAAAATCCCTTAACTTTCAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGC 5726  
Db 1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1140  
QY 5727 AATACCTCTCAGAC 5741  
Db 1141 AsnAsnLeuSerAsp 1145  
RESULT 4  
AD573447  
ID AD573447 standard; protein; 1145 AA.  
XX  
AC AD573447;  
XX  
DT 16-DEC-2004 (first entry)  
XX  
DE Swine retroviral pol protein.  
XX  
KW Swine retroviral protein; immunosuppressive; gene therapy; pol protein.  
XX  
OS Porcine endogenous retrovirus.  
XX  
PN US2004185435-A1.  
XX  
PD 23-SEP-2004.  
XX  
PF 26-NOV-2003; 2003US-00723552.

XX 14-DEC-1995; 95US-00572645.  
 PR 13-DEC-1996; 96US-00766528.  
 PR 14-SEP-2000; 2000US-00661858.  
 XX (GEO) GEN HOSPITAL CORP.  
 XX PA  
 XX PI Fishman JA;  
 XX WPI; 2004-689179/67.  
 DR N-PSDB; ADS73369.  
 XX  
 PT New porcine retroviral polypeptide encoded by a nucleic acid, useful in  
 PT evaluating an immunosuppressive treatment for the ability to activate a  
 PT retrovirus, such as an endogenous porcine retrovirus.  
 XX  
 PS Claim 8; Fig 3; 83pp; English.  
 XX  
 CC The present invention relates to the swine retroviral polypeptides and  
 CC their encoding nucleic acids. The methods and compositions of the present  
 CC invention are useful for screening a cell or tissue, e.g. a heart, lung,  
 CC liver, bone marrow, kidney, brain cells, neural tissue, pancreas and  
 CC intestinal tissue xenograft, for the presence or expression of a swine or  
 CC miniature swine retrovirus or retroviral sequence. The invention is also  
 CC useful in evaluating an immunosuppressive treatment for the ability to  
 CC activate a retrovirus such as an endogenous porcine retrovirus. The  
 CC invention is also useful in gene therapy. The present sequence is the  
 CC swine retroviral pol protein.  
 XX  
 SQ Sequence 1145 AA;  
 XX  
 Alignment Scores:  
 Pred. No.: 0 Length: 1145  
 Score: 6038.00 Matches: 1145  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 41.3% Indels: 0  
 DB: 8 Gaps: 0  
 US-10-723-552-3 (1-8132) x ADS73447 (1-1145)

QY	2787	TCTCCCTCGGAATAC	TCTCCCTCGGACTT	TAGAAAGCC	TGGGACTAAT	GACTATCGACCA	2846
DB	161	SerProTrpAsnThrProLeuLeuProValAlaGlySerProGlyThrAsnAspTyrArgPro	180				
QY	2847	GTACAGGACTTGAGAGAGGTCAATAAACGGGTGCAGGATATACACCAACAGTCCCGAAC	2906				
DB	181	ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn	200				
QY	2907	CCTTATAACCTCTTGCTGCTCTCCACCCCAACGGAGCTGGTATACAGTATTGGACTTA	2966				
DB	201	ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTrpTyrThrValLeuAspLeu	220				
QY	2967	AAGGATGCTCTCTCTGCTGAGATTACACCCCACTAGCAACCACTTTTGGCTTCGAA	3026				
DB	221	LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu	240				
QY	3027	TGGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTGCCCCCAAGGG	3086				
DB	241	TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly	260				
QY	3087	TTCAAGAACTCCCGGACCATCTTTGAGGAAGCCCTACACAGACCTGGCCAACTTCAGG	3146				
DB	261	PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg	280				
QY	3147	ATCCAAACCCCTCAGGTGACCCCTCCCTCCAGTACCTGGATCAGCTGCTTCTGGCGGGAGCC	3206				
DB	281	IleGlnHisProGlnValThrLeuLeuGlnTyrValAspLeuLeuLeuAlaGlyAla	300				
QY	3207	ACCAAAACAGGACTGTGTAGAAAGCCAGAGGCACTACTGCTGAAATGTCTGACCTAGGC	3266				
DB	301	ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuLeuSerAspLeuGly	320				
QY	3267	TACAGAGCTCTGCTAAGAAGCCCGAGATTGGCAGGAGAGTAACTATCTTTGGGGTAC	3326				
DB	321	TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr	340				
QY	3327	AGTTTCGGGACGGGACGCGATGGCTGACGGAGCGACGGAAGAAAAGTGTAGTCCAGATA	3386				
DB	341	SerLeuArgAspGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle	360				
QY	3387	CGGGCCCCAACCAACAGCCAAATGAGAGAGTGTGTGGGACAGCTGGATTTTGCAGA	3446				
DB	361	ProAlaProThrThrAlaLysGlnMetArgGluPheLeuGlyThrAlaGlyPheCysArg	380				
QY	3447	CTGTGATCCCGGGTTTGGACCTTAGCAGCCCACTTACCGCTACCCCAAGAAAA	3506				
DB	381	LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLys	400				
QY	3507	GGGGAATTCCTCGGGCTCTGAGCACCAAGAGCCATTTGATGCTATCAAAAAGGCCCTG	3566				
DB	401	GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu	420				
QY	3567	CTGAGCGCACCTGCTCTGGCCCTCTGACGTAACTAAACCCCTTTACCTTTATGTGGAT	3626				
DB	421	LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp	440				
QY	3627	GAGCTAAGGAGTAGCCCGGGAGTGTAAACCCCAACCTTAGACCATCGAGAGACCT	3686				
DB	441	GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro	460				
QY	3687	GTGCGCTACTCTGCAAGAGCTCGATCTGTGACCAAGTGTGTGGGCCATATGCTCGAAG	3746				
DB	461	ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLys	480				
QY	3747	GCTATCCGAGCTGTGGCCATCTGGTCAAGGACGCTGACAAATGTGCTTTGGGACAGAAT	3806				
DB	481	AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn	500				
QY	3807	ATNACTGTAAATAGCCCCCATGCTTGGAGAACATGTTCCGCGAGCCCCCAGACCGATGG	3866				
DB	501	IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp	520				
QY	3867	ATGACCAACGCGCGCATGATGATCAAAAGCCTGCTTCTCACAGAGAGGTGACGTTTC	3926				

```

Db      521 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 540
QY      3927 GCTCCACGACCGCTCTCAACCCCTGCGACCTCTTCTGCTCAGAGACTGATGAACAGTG 3986
Db      541 AlaProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 560
QY      3987 ACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGTCCGCAAGGACCTTACAGAC 4046
Db      561 ThrHisAspCysHisGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 580
QY      4047 ATACCGCTGACTGAGAAAGTCTAACTGCTTCACTGACGGAAGCAGCTATGCTGGAA 4106
Db      581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 600
QY      4107 GGTAAGAGGATGGCTGGGGCGGGTGTGCGACGGGACCGCCACGATCTGGGCGACGAGC 4166
Db      601 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 620
QY      4167 CTGCGGAGAGAACTTCAGCACAAAGGCTGAGCTCATGCCCTCAGCGCAAGCTTTGCGG 4226
Db      621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640
QY      4227 CTGCGCGAAGGAAATCCATAACATTTATACGACAGAGGATGCTTTCGACTGCA 4286
Db      641 LeuAlaGlnGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 660
QY      4287 CAGCTACATGGGCGCATCTATAACAAGGGGTGTCTTACCTCAGCAGGAGGGAAATA 4346
Db      661 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680
QY      4347 AAGAAACAAGAGAAATTTAAGCCCTATTAGAGCCGTACATTTACCAAAAGGCTAGCT 4406
Db      681 LysAsnLysGluLeuLeuSerLeuLeuGluAlaValHisLeuProLysArgLeuAla 700
QY      4407 ATTATACACTCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGAAACAGATG 4466
Db      701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720
QY      4467 GCTGACCGGGTTCCTCAAGCAGGCGCCCGGGTGTAACTCTGCTTATATAGAAATG 4526
Db      721 AlaAspArgValAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProIleIleGluMet 740
QY      4527 CCAAAAGCCCAAGACCCAGACAGACAGTACACCTAGAGACTGGCAAGAGATAAAAG 4586
Db      741 ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760
QY      4587 ATAGACAGTTCTCTGAGACTCGGAAGGACCTGCTATACCTCAGATGGGAAGGAATC 4646
Db      761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 780
QY      4647 CTGCCCCACAAGAAGGGTTAGAAATATGTCCAACAGATACATCGTCTAACCCACCTPAGA 4706
Db      781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800
QY      4707 ACTAAACACTGCGAGTGTGTGAGAACATCCCTTATCATGTCTCTGAGGCTACCAGA 4766
Db      801 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820
QY      4767 GTGGCTGACTCGGTGTGCTCAACATTTGTGTGCTGCGCTGACCTGGTAACTCACTCTCC 4826
Db      821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840
QY      4827 AGAATCCCTCCAGGGAAGAGACTTAAGGGAAGCCACCGAGGCGCTCACCTGGGAAGTGGAC 4886
Db      841 ArgMetProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860
QY      4887 TTCCTGAGGTAAAGCCGCTAAATACGGAACAAATACCTATTGTGTTTTGTAGACACC 4946
Db      861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880
QY      4947 TTTTCAGGATGGTAGAGCTTATCTCTACTAGAAAGAGACTTCAACCGTGTGGCTAAA 5006

```

```

Db      881 PheSerGlyTrpValGluAlaTyrProThrLysGluThrSerThrValValAlaLys 900
QY      5007 AAAATACTGGAAGAAATTTTTTCCAGATTTGGAAATACCTAAAGCTAATAGGCTCAGACAAT 5066
Db      901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920
QY      5067 GFTCCAGCTTTTGTTCCTCCAGGTAACTCAGGGACTGGCCCAAGATATTGGGGATTGATTGG 5126
Db      921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940
QY      5127 AAATCTCATGTGCTATACAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAAC 5186
Db      941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960
QY      5187 ATTAAAGAGACCTTACTAAATTTGACCGCGAGACTGGCGTTAATGATTGGATAGCTCTC 5246
Db      961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTrpIleAlaLeu 980
QY      5247 CTGCCCTTTGTGCTTTTAGGTTAGGAAACACCCCTGGACAGTTTGGGCTGACCCCTTAT 5306
Db      981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000
QY      5307 GAATTACTCTACGGGGGACCCCCCTTGGTAGAATTTGCTTCTGTCTGTACTAGTCTGCTGAC 5366
Db      1001 GluLeuLeuTyrGlyGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1020
QY      5367 GTGCTGCTTTCCAGGCTTTGTTCTTAGGCTCAAGGCACTTGAGTGGGTGAGACAACGA 5426
Db      1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040
QY      5427 GCGTGGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGAGACTTGCAGATCCACATCGT 5486
Db      1041 AlaTrpArgGlnLeuArgGluAlaTyrSerGlyGlyGlyAspLeuGlnIleProHisArg 1060
QY      5487 TTCCAAGTGGGAGATTCACTAGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGG 5546
Db      1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080
QY      5547 TCGAAGGGCCCTTATCTGCTACTTTTACCACCAACCGCTGTGAAAGTCCGAAGGAATC 5606
Db      1081 TrpLysGlyProTyrLeuValLeuLeuThrProThrAlaValLysValGluGlyIle 1100
QY      5607 TCCACTCGATCCATCCATCCCACTTAAACCGCGCCACCTCCCGATTTCGGGGTGGAAA 5666
Db      1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTrpLys 1120
QY      5667 GCGCAAAAGACTGAATAATCCCTTAAAGCTTCGCTCCATCGCGTGGTTCTTACTCTGTC 5726
Db      1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1140
QY      5727 AATAACCTCTCAGAC 5741
Db      1141 AsnAsnLeuSerAsp 1145

```

RESULT 5

AAB10044

ID AAB10044 standard; protein; 1737 AA.

XX AAB10044;

XX 02-NOV-2000 (first entry)

XX MMLV gag-pol protein.

XX Glycoprotein; gag gene; pol gene; GP-1; GP-2; anti-HIV; cytostatic;

XX gene therapy; treatment; infectious disease; HIV; AIDS; neoplasm;

XX carcinoma; melanoma.

XX Moloney murine leukemia virus.

XX EP1006196-A2.

XX 07-JUN-2000.

XX 25-NOV-1999; 99EP-00250415.  
 XX 26-NOV-1998; 98DE-01056463.  
 PR (PETT-) PETTE INST HEINRICH.  
 XX Von Laer MD;  
 XX WPI; 2000-378268/33.  
 XX New retroviral packing cell useful as pharmaceutical carrier in gene  
 PT therapy for treatment of HIV and neoplasms, comprises retroviral genes  
 PT and glycoproteins.  
 XX Disclosure; Page 32-36; 69pp; German.  
 XX This invention describes a novel retroviral packing cell (I), comprising  
 CC the retroviral genes gag, pol and glycoproteins gp-1 and gp-2 of the LCMV  
 CC coding gene gp, or a part of these. The products of the invention have  
 CC anti-HIV and cytostatic activity and can be used for gene therapy. (I) is  
 CC useful for in vitro infection of cells, especially hematopoietic stem  
 CC cells, for expression of transgenes in cells and as a pharmaceutical  
 CC carrier for gene therapy. (I) is therefore useful in the treatment of  
 CC infectious diseases (HIV or AIDS), neoplasms (carcinomas), melanomas and  
 CC other diseases. This sequence represents the Moloney murine leukemia  
 CC virus (MMLV) gag-pol protein which is described in the method of the  
 CC invention  
 XX Sequence 1737 AA;  
 SQ

Alignment Scores:  
 Pred. No.: 0 Length: 1737  
 Score: 5338.00 Matches: 1035  
 Percent Similarity: 71.6% Conservative: 239  
 Best Local Similarity: 58.2% Mismatches: 395  
 Query Match: 36.5% Indels: 110  
 DB: 3 Gaps: 26

US-10-723-552-3 (1-8132) x AAB10044 (1-1737)

QY 585 ATGGGACAGCGGTGACGACCCCTCTTAGTTGACTCTCGACCATGGACTGAAGTAA 644  
 DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20  
 QY 645 TCAGGGCTCATATTTGTCAGTTAGCTTACAGGACCTTGGCAGACTTCTGTGTC 704  
 DB 21 ArgIleAlaHisAsnGlnSerValAspValLysLysArgArgTrpValThrPheCysSer 40  
 QY 705 TCTGATGGCCGACATTTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATCTTGAGATT 764  
 DB 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60  
 QY 765 ATCCTGGCTCTTAAAGCAGTTATTTTTCAGACTGGACCCCGGCTCTCATCCCGATCAGGAG 824  
 DB 61 IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80  
 QY 825 CCCTATATCTTACGTGGCAAGATTGTCGACAGAGATCTCCGCCATCGGTTAAACCATGG 884  
 DB 81 ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProTrpValLysProPhe 100  
 QY 885 CTGAATAAGCCCAAGAACCCAGGTCCCGAATTCCTGCTCTTGGAGAGAAACCAACAC 944  
 DB 101 ValHis---ProLysProProProProLeuProProSerAlaProSerLeuProLeuGlu 119  
 QY 945 TCGGCTGAATAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGACCA---CCG 1001  
 DB 120 ProProArgSerThrProProArgSerSerLeuTyrProAlaLeuThrProSerLeuGly 139  
 QY 1002 GCTTGGCCGGAACCCCA-----TCT 1022  
 DB 140 AlaLysProLysProGlnValLeuSerAspSerGlyGlyProLeuIleAspLeuLeuThr 159

QY 1023 GTTCCCCCACCCTTATCTGCGACAGGGTCCGCGAGGGAGCCCTTTGCCCTCTCT--- 1079  
 DB 160 GluAspProProTyr-----ArgAspProArgProProSer 173  
 QY 1080 -----GGAGCTCGGCGGTGAGGACCT----- 1103  
 DB 174 AspArgAspGlyAsnGlyGlyGluAlaThrProAlaGlyGluAlaProAspProSerPro 193  
 QY 1104 ---GCTGAGGAGACTCGGAGCGGCGGCCACCCCGGAGCGGACAGAGATCGCG 1160  
 DB 194 MetAlaSerArgLeuArgGlyArgGluProProValAlaAspSerThrThrSerGln 213  
 QY 1161 ACATTACCCTCGCACGTACGCGCTCCACACACGCGGGGCGCAATGCGACCCCTCCAG 1220  
 DB 214 AlaPheProLeuArgAlaGlyGly-----AsnGlyGln-----LeuGln 236  
 QY 1221 TATTGGCCCTTTCTTCAGATCTCTATAATTGAAAACTAACCATCCCTCTCTCG 1280  
 DB 227 TyrTrpProPheSerSerAspLeuTyrAsnTrpLysAsnAsnAsnProSerPheSer 246  
 QY 1281 GAGGATCCCAACGCTCAGCGGTGGTGGAGTCCCTTATGTTCTCTCACCAGCCTACT 1340  
 DB 247 GluAspProGlyLysLeuThrAlaLeuIleGluSerValLeuIleThrHisGlnProThr 266  
 QY 1341 TGGGATGATTGTCAACAGCTGTGCAGACACTCTTCAACACGAGGAGCGAGAGAAAT 1400  
 DB 267 TrpAspAspCysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGluGlyLysGlnArgVal 286  
 QY 1401 CTATTAGAGGTAGAAAAAATGTTCTGGGCGCGAGCGGCGGACCCACGCGGTGCAAAAT 1460  
 DB 287 LeuLeuGluAlaArgLysAlaValArgLysAspAspGlyArgProThrGlnLeuProAsn 306  
 QY 1461 GAGATTGACATGGATTTCCTTAATCTCGCCCGGTGGGACTACAAACGCTGAAGGT 1520  
 DB 307 GluValAspAlaAlaPheProLeuGluArgProAspTrpAspTrpThrGlnAlaGly 326  
 QY 1521 AGGGAGAGCTTGAAAAATCTATCGCCAGGCTCTGGTGGCGGCTCTCGGCGCGCTCAAGA 1580  
 DB 327 ArgAsnHisLeuValHisTyrArgGlnLeuLeuAlaGlyLeuGlnAsnAlaGlyArg 346  
 QY 1581 CGGCCCATTAATTTGGCTAAGTAAGTAAGTATGTCAGGAGCCGGAATGAACCCCTCT 1640  
 DB 347 SerProThrAsnLeuAlaLysValLysGlyIleThrGlnGlyProAsnGluSerProSer 366  
 QY 1641 GTTTTCTTGGAGGCTCTTGAAGCCTTCAGCGGTACACCCCTTTTATCCACCTCA 1700  
 DB 367 AlaPheLeuGluArgLeuLysGluAlaTyrArgTyrThrProTyrAspProGluAsp 386  
 QY 1701 GAGGCCCAAAAGCCTCAGTGGCTTTGGCCTTTATAGGACAGTCAGCTTGGATATTAGA 1760  
 DB 387 ProGlyGlnGluThrAsnValSerMetSerPheIleTrpGlnSerAlaProAspIleGly 406  
 QY 1761 AGAAGCTTCAGACTGGAAGGTTACAGGAGGCTGAGTTACGTATCTAGTGAAGGAG 1820  
 DB 407 ArgLysLeuGluArgLeuGluAspLeuLysAsnLysThrLeuGlyAspLeuValArgGlu 426  
 QY 1821 GCAGAAAGTATATTACAAAAGGAGACAGAGAAGAAAGGCAACAAAGAAAAGAGA 1880  
 DB 427 AlaGluLysIlePheAsnLysArgGluThrProGluGluArgGluArgGly 446  
 QY 1881 GAAAGAGAGAAAGGAGGAGAAAGA-----CGT 1907  
 DB 447 GluThrGluGluLysGluGluArgArgThrGluAspGluGlnLysGluLysGluArg 466  
 QY 1908 ATAATAACGCGACAGAGAAGAAATTTGACTAAGATCTTGGCTGCAGTGGTTGAGGAAAAGC 1967  
 DB 467 AspArgArgHisArgGluMetSerLysLeuLeuAlaThrValValSerGlyGlnLys 486  
 QY 1968 AATACGGAAGAGAGAGAGATTATTAGGAAAATTTAGTTCAGGCCCTAGACAGTCAGGGAAC 2027  
 DB 487 Gln-----AspArgGlnGlyGlyGlu 493  
 QY 2028 CTGGGCAATAGGACCCCTCTCGACAAAGGACCAATGTGCATATTGTAAAGAAAGAGGACAC 2087



QY 4221 TTGCGGCTGCGCAAGGGAATCCATTAACATTTATACGACAGCAGGTATGCTTTGCG 4280  
 Db 1228 LeuLysMetAlaGluGlyLysLysLeuAsnValThrAspSerArgTyrAlaPheAla 1247  
 QY 4281 ACTGCACACGTACATGGGGCCATCTATAACAAAGGGGTGCTTACCTCAGCAGGAGG 4340  
 Db 1248 ThrAlaHisIleHisGlyGluIleTyrArgArgGlyLeuLeuThrSerGluGlyLys 1267  
 QY 4341 GAAATAAGAACAAAGAGGAAATCTAAGCCTATTAGAGCCCTACATTTACCAAAAGG 4400  
 Db 1268 GluIleLysAsnLysAspGluIleLeuAlaLeuLysAlaLeuPheLeuProLysArg 1287  
 QY 4401 CTAGCTATTATACACTGCTGACATCAGAAAGCTAAAGATCTCATATCCAGAGAAAC 4460  
 Db 1288 LeuSerIleIleHisCysProGlyHisGlnLysGlyHisSerAlaGluAlaArgGlyAsn 1307  
 QY 4461 CAGATGGCTGACCGGGTTGCCAAGCAGGAGCCAGCGGTGTTAACCTTCTGCTATAATA 4520  
 Db 1308 ArgMetAlaAspGlnAlaAlaArgLysAlaAla-----IleThr 1320  
 QY 4521 GAAATGCCAAAGCC----- 4547  
 Db 1321 GluThrProAspThrSerThrLeuLysLeuGluAsnSerSerProTyrThrSerGluHis 1340  
 QY 4548 CGACAGTACACCTAGACACTGCGAAGAGATATAAAAGATAGACCAG--TTCTCTGAG 4604  
 Db 1341 PheHieTyrThrValThrAspIleLysAspLeuThrLysLeuGlyAlaIleTyrAspLys 1360  
 QY 4605 ACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAGAAATCTGCCCAAAAGAGGG 4664  
 Db 1361 ThrLysLysTyrTrpValTyr-----GlnGlyLysProValMetProAspGlnPheThr 1378  
 QY 4665 TTAGAATATGTCCAACAGATACATCGTCTAACCCACTAGGAATAACACCTGACGACG 4724  
 Db 1379 PheGluLeuLeuAspPheLeuHisGlnLeuThrHisLeuSerPheSerLysMetLysAla 1398  
 QY 4725 TTGGTCAGACA-----TCCCTTATCATGTTCTGAGGCTACGAGGTGGCTGACTCG 4778  
 Db 1399 LeuLeuGluArgSerHisSerProTyrTyrMetLeuAsnArgAspArgThrLeuLysAsn 1418  
 QY 4779 GTGGTCAAAACATTGTGTCCTGCCAGCTGTTAAATGCTTAATCTTCCAGAATGCCCTCA 4838  
 Db 1419 IleThrGluThrCysLysAlaCysAlaGlnValAsnAlaSerLysSerAlaValLysGln 1438  
 QY 4839 GGAAGAGACTAAGGGAAGCCACCCAGGCGCTCACTGGGAATGCACTTCACTGAGGTA 4898  
 Db 1439 GlyThrArgValArgGlyHisArgProGlyThrHisTrpGluIleAspPheThrGluIle 1458  
 QY 4899 AAGCCGGCTAAATACGMAACAATACCTATTGTTTGTAGACACCTTTTCAGGATGG 4958  
 Db 1459 LysProGlyLeuTyrGlyTyrLysTyrLeuLeuValPheIleAspThrPheSerGlyTrp 1478  
 QY 4959 GTAGAGCTTATCTACTAAGAAGAGACTTCAACCGTGGTGGCTAAAAAATACTGGAA 5018  
 Db 1479 IleGluAlaPheProThrLysLysGluThrAlaLysValValThrLysLysLeuLeuGlu 1498  
 QY 5019 GAAATTTTCCAAGATTGGAATACCTAAGGTAATAGGTCAGACAATGTCGACGCTTTT 5078  
 Db 1499 GluIlePheProArgPheGlyMetProGlnValLeuGlyThrAspAsnGlyProAlaPhe 1518  
 QY 5079 GTTGGCCAGTAACTCAGGACTGGCCAGATATTGGGATTCATTGGAACTGCAATTGT 5138  
 Db 1519 ValSerLysValSerGlnThrValAlaAspLeuGlyLysAspTrpLysLeuHisCys 1538  
 QY 5139 GCATACAGACCCCAAGCTCAGACAGGTAGAGAGGATGAATAGAACCATTAAGAGACC 5198  
 Db 1539 AlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThr 1558  
 QY 5199 CTTACTAAATTTGACCGGAGACTGGCGTTAATGATTGGATAGCTCTCTGCGCCTTTG 5258  
 Db 1559 LeuThrLysLeuThrLeuAlaThrGlySerArgAspTrpValLeuLeuLeuProLeuAla 1578

QY 5259 CTTTTTAGGTTAGGAACACCCCTGGACAGTTTGGCTGACCCCTATGAATTACTCTAC 5318  
 Db 1579 LeuTyrArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGluIleLeuTyr 1598  
 QY 5319 GGGGACACCCCCCATTTGGTAGAAATTGCTTCTGTACATAGTGTCTGACGTGCTTTCC 5378  
 Db 1599 GlyAlaProProLeuValAsnPheProAspProAspMetThrArgValThrAsnSer 1618  
 QY 5379 CAGCCTTTGTTCTTAGGCTCAAGGACTTGTAGTGGTGAGACAACGAGCGTGGAGCAA 5438  
 Db 1619 ProSerLeuGlnAlaHisLeuGlnAlaLeuTyrLeuValGlnHisGluValTrpArgPro 1638  
 QY 5439 CTCGGGAGCGCTACTCAGGAGGAGAGACTTGCAG--ATCCACATCGTTTCCAAGTG 5495  
 Db 1639 LeuAlaAlaAlaTyrGlnGlnLeuAspArgProValValProHisProTyrArgVal 1658  
 QY 5496 GGAGATTCACTTACGTTAGACCCACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGC 5555  
 Db 1659 GlyAspThrValTrpValArgHisGlnThrLysAsnLeuGluProArgTrpLysGly 1678  
 QY 5556 CTTATCTCTGTTTGTGACCAACCGCTGTGAAGTCAAGGAATCTCCACCTGG 5615  
 Db 1679 ProTyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleAlaAlaTrp 1698  
 QY 5616 ATCCATGTCATCCACGTTAAACCGCGCCACCT-----CCCGATTGCG-----GGG 5660  
 Db 1699 IleHisAlaAlaHisValLysAlaAlaAspProGlyGlyProSerSerArgLeuThr 1718  
 QY 5661 TGGAAAGCCGAAAGACTGAAATCCCTTAAGCTTTCGCTCCATCGCTGGTTCCT 5717  
 Db 1719 TrpArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGluAlaPro 1737  
 RESULT 6  
 AAR94427  
 ID AAR94427 standard; protein; 1784 AA.  
 AC AAR94427;  
 XX 16-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 11-JUN-1996 (first entry)  
 XX FeLV F6A provirus clone 61E encoded nucleocapsid protein.  
 XX FeLV; retrovirus; vaccine; AIDS; disease model; immunodeficiency;  
 KW viraemia; leukaemia; therapy; nucleocapsid.  
 OS Feline immunodeficiency virus; subtype A.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 581 /notes "codon 581 in encoding sequence is TAG"  
 FT  
 XX EP699758-A1.  
 PN  
 XX 06-MAR-1996.  
 PD  
 XX 12-DEC-1989; 95EP-00100871.  
 PF  
 XX 13-DEC-1988; 88US-00284139.  
 PR  
 XX (HARD ) UNIV HARVARD.  
 PA (COLS ) UNIV COLORADO STATE RES FOUND.  
 XX  
 PI Hoover EA, Mullins JI;  
 XX WPI; 1996-180826/19.  
 DR N-PSDB; AAT13265.  
 DR  
 XX Inactivated FeLV-A sub-type isolates for use in disease models and  
 PT vaccines - can be used to study prophylaxis and therapy of related immuno  
 PT -deficiencies in other species, e.g. human.  
 XX





Db 602 oValThrPheLeuValAspThrGlyAlaGlnHisSerValLeuThrArgProAspGlyPr 622  
QY 2282 ACTAAAGATAAAATCCCTGGGTGATGGTGCACACAGGCCAACACAGTATCATGGAC 2341  
Db 622 oLeuSerAspArgThrAlaLeuValGlnGlyAlaThrGlySerLysAsnTyrArgTrpTh 642  
QY 2342 TACCCTGAAGAACAGTTGACTGTGGAGTGGGACGGGTAAACCCACTCGTTTCTGTGTCATACC 2401  
Db 642 rThrAspArgValGlnLeuAlaThrGlyLysValThrHisSerPheLeuTyrValPr 662  
QY 2402 TGAGTCCCCAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATGGGACGACAAATTC 2461  
Db 662 oGluCysProTyrProLeuLeuGlyArgAspLeuLeuThrLysLeuAlaGlnIleHi 682  
QY 2462 TTTTGA---CAAGGGAACACAGAGTGTCTGCAATAACAACCTATCACTGTGTGAC 2518  
Db 682 sPheThrGlyGluGlyAlaAsnValValGlyProArgGlyLeuProLeuGlnValLeuTh 702  
QY 2519 CTCCCAATTAGATGACGAATATCGACTATATCTCTCCCTAGTAAAGCCTGATCAAAATAT 2578  
Db 702 rLeuGlnLeuGluGluGluTyrArgLeuPheGluProGluSerThrGlnLysGlnGluMe 722  
QY 2579 ACAATTCTGTTGGACAGTTTCCCAAGCCTCGGCAGAACCGCAGGATGGGTTGGC 2638  
Db 722 tAspIleTrpLeuLysAsnPheProGlnAlaTrpAlaGluThrGlyGlyMetGlyMetal 742  
QY 2639 AAGCAAGTTCCCCACAGATTATTCACCTGAAGGCCAGTGCCACACAGTGTCACTGAC 2698  
Db 742 aHisCysGlnAlaProValLeuIleGlnLeuLysAlaThrAlaThrProIleSerIleAr 762  
QY 2699 ACAGTACCCCTTCAGTAAAGAACTCAAGAAGAAATTCGGCCCGCATGTCCAAAGATTAT 2758  
Db 762 gGlnTyrProMetProHisGluAlaTyrGlnGlyIleLysProHisIleArgMetLe 782  
QY 2759 CCAACAGGGCATCTAGTTCTCTGTCCATCTCCCTGGAATPACTCCCTGTCTACCGTTAG 2818  
Db 782 uAspGlnGlyIleLeuLysProCysGlnSerProTrpAsnThrProLeuLeuProVally 802  
QY 2819 AAGCCCTGGGACTAATGACTATCGACCACTACAGGACTTCAGAGAGGTCAATAACGGGT 2878  
Db 802 sLysProGlyThrLysAspTyrArgProValGlnAspLeuArgGluValAsnLysArgVa 822  
QY 2879 GCAGGATATACCCAAACAGTCCGAAACCTTATACTCTGTGTCTCCACCCCA 2938  
Db 822 lGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerThrLeuProProSe 842  
QY 2939 ACGGAGCTGTATACAGTATTGGACTTAAAGTAGCCTCTCTCTGCTGAGATTACACCC 2998  
Db 842 rHisProTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisSe 862  
QY 2999 CACTAGCCAAACCACTTTTTCCTTCGAAATGGAGAGATCCAGGTACCGGAAGAACCGGCA 3058  
Db 862 rGluSerGlnLeuLeuPheAlaPheGluTrpArgAspProGluIleGlyLeuSerGlyGl 882  
QY 3059 GCTCACTGGACCCGCTGCCCAAGGGTCAAGAACTCCCGCCACCATCTTTCACGAAGC 3118  
Db 882 nLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAspGluAl 902  
QY 3119 CCTACACAGACCTGGCCAACTTCAGGATCCAAACACCCCTCAGGTGACCCCTCTCCAGTA 3178  
Db 902 aLeuHisSerAspLeuAlaAspPheArgValArgTyrProAlaLeuValLeuLeuGlnTy 922  
QY 3179 CGTGGATGACCTGCTTCTGGCGGAGCCACCAACAGGACTGCTTAGAGGCGACGAAGGC 3238  
Db 922 rValAspAspLeuLeuAlaAlaAlaThrArgThrGluCysLeuGluGlyThrLysAl 942  
QY 3239 ACTACTGCTGAATTGCTGACCTAGGCTACAGAGCTCTGCTAAAGAGGCCAGATTTCG 3298  
Db 942 aLeuLeuGluThrLeuLeuGlyAsnLysGlyTyrArgAlaSerAlaLysLysAlaGlnIleCy 962  
QY 3299 CAGGAGAGGTTAACTACTTGGGGTACAGTTTTCGGGACGGGACGAGTGGCTGACGGA 3358  
Db 962 sLeuGlnGluValThrTyrLeuGlyTyrSerLeuLysAspGlyGlnArgTrpLeuThrly 982

QY 3359 GGCACGGAAGAAACCTGTAGTCCAGATACCGGCCCAACACACAGCCAAACAAATGAGAGA 3418  
Db 982 sAlaArGlySGluAlaIleLeuSerIleProValProLysAsnProArgGlnValArgGl 1002  
QY 3419 GTTTTGGGGACAGCTGGATTTTTCAGACTGTGGATCCCGGGGTTTTCGCACTTAGCAGC 3478  
Db 1002 uPheLeuGlyThrAlaGlyTyrCysArgLeuTrpIleProGlyPheAlaGluLeuAlaAl 1022  
QY 3479 CCCACTCTACCCGCTAACCAAGAAAGGGGAATTTCTCTGGGCTCTCGAGCACCAGAA 3538  
Db 1022 aProLeuTyrProLeuThrArgProGlyThrLeuPheGlnTrpGlyThrGluGlnGlnLe 1042  
QY 3539 GGCATTGTGATGCTATCAAAAGSCCTGCTGAGCGCACCTGCTCTCGCCCTCCCTCAGCT 3598  
Db 1042 uAlaPheGluAsnIleArgLysAlaLeuLeuSerSerProAlaLeuGlyLeuProAspI 1062  
QY 3599 AACTAAACCTTTTACCCTTTATGTGATGAGCGTAAAGGAGTAGCCCGGGAGCTTTTAAC 3658  
Db 1062 eThrLysProPheGluLeuPheIleAspGluAsnSerGlyPheAlaLysGlyValLeuVa 1082  
QY 3659 CCAACCCCTAGGACCATGAGAGAAGCTGTGCTGCTACCTGTCAAGAAGCTCGATCCTGT 3718  
Db 1082 lGlnLysLeuGlyProTrpLysArgProValAlaTyrLeuSerLysLysLeuAspThrVa 1102  
QY 3719 AGCCAGTGTGGCCCATATGCTGAAGCTATCGCAGCTGTGGCCATCTACTGCTCAAGGA 3778  
Db 1102 lAlaSerGlyTrpProCysLeuArgMetValAlaAlaIleAlaIleLeuValLysAs 1122  
QY 3779 CGCTGCAAAATGACTTTTGGGACAGAAATATACTGTAATAGCCCCCATCATGTCATGGAGAA 3838  
Db 1122 pAlaGlyLysLeuThrLeuGlyGlnProLeuThrIleLeuThrSerHisProValGluAl 1142  
QY 3839 CATCGTTGGCAGCCCCCAGACCGATGATGATGACCAACGCCCGCATGATCAACACTATCAAAG 3898  
Db 1142 aLeuValArgGlnProProAsnLysTrpLeuSerAsnAlaArgMetThrHisTyrGlnAl 1162  
QY 3899 CTGCTTCTC---ACAGAGAGGTCACGTTTCGCTCCACAGCCGCTCTCAACCTCGCCAC 3955  
Db 1162 aMetLeuLeuAspAlaGluArgValHisPheGlyProThrValSerLeuAsnProAlaTh 1182  
QY 3956 TCTTCTGCTCAGAGAGACTGATGAACCA-----GTGACTCATGATTGCCATCAACTATT 4009  
Db 1182 rLeuLeuProLeuProSerGlyLysProArgLeuSerProAsp----- 1197  
QY 4010 GATTGAGGAGACTGGGGTCCGCAAGCACTTACAGACATACCGCTGCTGAGGAGAGTGCT 4069  
Db 1198 -LeuAlaGluThrMetAlaGlnThrAspLeuThrAspGlnProLeuProAspAlaAspLe 1217  
QY 4070 AACCTGGTTCACTGACGGAAGCAGCTATGTTGGTGAAGGTAAAGAGATGGCTGGCGGCGC 4129  
Db 1217 uThrTrpTyrThrAspGlySerSerPheIleArgAsnGlyGluArgLysAlaGlyAlaAl 1237  
QY 4130 GSTGTGGACGGGACCCGACCATCTGGGCGCAGCAGCTGCGCGAAGGAAGCACTTCAGCACA 4189  
Db 1237 aValThrThrGluSerGluValIleTrpAlaAlaSerLeuProProGlyThrSerAlaGl 1257  
QY 4190 AAAGCTGAGCTCATGGCCCTCAGCAAGCTTTTGGCGCTGGCCGGAAGGAAATCCATAAA 4249  
Db 1257 nArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaLysGlyLysLeuTh 1277  
QY 4250 CATTTATACGACAGCAGGTATGCTTTCGCTGCTGACGTACACGTACATGCGGCCCATCTATAA 4309  
Db 1277 rValTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisGlyGluIleTyrAr 1297  
QY 4310 ACAAGGGGGTTCCTTACCTCCTCAGCAGGAGGGAATAAAGAAACAAAGAGGAAATTCAGAG 4369  
Db 1297 gArgArgGlyLeuLeuThrSerGluGlyLysGluIleLysAsnLysAsnGluIleLeuAl 1317  
QY 4370 CCTATTAGAGCCGTTACATTACCAAAAAAGCTAGCTATTATACACTGTCTCGCATCA 4429  
Db 1317 aLeuLeuGluAlaLeuPheLeuProLysArgLeuSerIleHisCysProGlyHisGl 1337

QY	4430	GAAGCTAAGATCTATATCCAGAGGAAACGAGATGCTCAACGGGTGTCACGACGCG	4489
DB	1337	nylyGlyAspSerProGlnAlaLysGlyAsnArgLeuAlaAspAspThrAlaLysAl	1357
QY	4490	AGCC-----CAGGGTGTTAACTCTTCGCT-----ATAATAGAAATGCCAA	4531
DB	1357	aAlaThrGluThrGlnSerSerLeuThrIleLeuProThrGluLeuIleGluGlyProLy	1377
QY	4532	AGCCCCAGAACCCAGACACAGTACACCTAGAAAGACTGGCNAGAGATAAANAAGATGA	4591
DB	1377	s-----ArgProProTyrGluTyrAspAspSerAspLeuAspLeuValGlnLysLeuGl	1395
QY	4592	CGAGTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATCGGAAGAAATCTCGCC	4651
DB	1395	uAlaHisTyrGluProLysAArgGlyThr---TrrpLutyArgGlyLysThrIleMetPr	1414
QY	4652	CCACAAAGAAGGTTAGAAATATGTCACACAGATACATCGTCTACCCACCTTAGGACTAA	4711
DB	1414	oGluLysTyrAlaLysGluLeuIleSerHisLeuHisLysLeuThrHisLeuSerAlaAr	1434
QY	4712	ACACCTCAGCAGGTGGTGCACAACATCC-----CCTTATCATGTCT	4753
DB	1434	gLyseMetLysThrLeuLeuGluAArgGluGluThrGlyPheTyrLeuProAsnArgAspLe	1454
QY	4754	GAGGCTACCGAGGTGGCTGACTCGGTGGTCAAAATGTGTGCCCTGCCAGCTGGTAA	4813
DB	1454	uHisLeuArgGlnValThrGluSer-----CysArgAlaCysAlaGlnIleAs	1470
QY	4814	TGCTAATCCTCCAGATGCCCTCCAGGAGAGAGACTAAGGGAGAGCCACCAGCGCTCA	4873
DB	1470	naLsGlyLysIleLysPheGlyProAspValArgAlaArgGlyArgArgProGlyThrHi	1490
QY	4874	CTGGGAAGTGCATCTCAGGAGTAAAGCCGGCTAAATACGGAACAAATACCTATTGGT	4933
DB	1490	sTrrpLyuValAspPheThrGluIleLysProGlyMetTyrGlyTyrLysTyrLeuLeuVa	1510
QY	4934	TTTTGTAGACACCTTTTCAGAGTGGGTAGAGCTTATCCTACTAAGAAAGAGACTCAAC	4993
DB	1510	lPheIleAspThrPheSerGlyTrrpAlaGluAlaTyrProAlaLysHisGluThrAlaLy	1530
QY	4994	CGTGGTGGCTAAAAAATACTGGAGAAATTTTCCAGATTTTGGATACCTAAGGTAA	5053
DB	1530	sValValAlaLysLysLeuGluGluIlePheProArgTyrGlyIleProGlnValle	1550
QY	5054	AGGCTCAGACAAATGGTCAGCTTTTGTGCCAGGTAAAGTCAGGACTGCCCAAGATATT	5113
DB	1550	uGlySerAspAsnGlyProAlaPheIleSerGlnValSerGlnSerValAlaThrLeuLe	1570
QY	5114	GGGATTGATTGGAACTGCATTGTGCATACAGACCCCAAGCTCAGGACAGGTAGAGAG	5173
DB	1570	uGlyLeAsnTrrpLysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluAr	1590
QY	5174	GATCAATAGAACCATTAAGAGACCCCTTACTAAATTGACCCGGAGAGCTGGCGTTAATGA	5233
DB	1590	gMetAsnArgSerIleLysGluThrLeuThrLysLeuThrLeuGluThrGlySerLysAs	1610
QY	5234	TTTGATAGTCTCTCTGCGCTTTGTGCTTTTGTAGGTTAGGAACACCCCTTGACAGTTTG	5293
DB	1610	pTrrpValLeuLeuLeuProLeuValLeuTyrArgValArgAsnThrProGlyProHisGl	1630
QY	5294	GCTGACCCCTATGAATTACTCTACGGGGAGCCCCCCCATTGGTAGAAATTCCTCTGT	5353
DB	1630	yLeuThrProPheGluIleLeuTyrGlyAlaProProProMetAlaHisPhePheAspTh	1650
QY	5354	ACATAGTGTGACGTGTGCTTTTCCAGCCCTTTGTTCTTAGGCTCAAGGCATTTGAGTG	5413
DB	1650	rAspIleSerThrPheAlaThrSerProThrMetGlnAlaHisLeuArgAlaLeuGlnLe	1670
QY	5414	GGTCAGACAAAGCGGTGGAGGCAACTCCCGGAGGCTACTCAGGAGGAGGACTTGCA	5473
DB	1670	uValGlnGluIleGlnArgProLeuAlaAlaTyrArgGluLysLeuGluThrPr	1690
QY	5474	G-----ATCCACATCGTTTCCAAGTGGGAGATTTCAGTCTACGTTAGACGCCACCGTGCAGG	5530

Db	1690	oValProHisProPhelysProGlyAspSerValTrpValArgHisGlnThrLy	1710		
Qy	5531	AAACCTCGAGACTCGGTGGAAGGCCCTTATCTCGTACTTTTGACCAACCAACGGCTGT	5590		
Db	1710	sAsnLeuGluProArgTrpYsGlyProHisIleValLeuLeuThrThrProThrAlaLe	1730		
Qy	5591	GAAGTTCGAAGGAATCTCCACTCGATCCATGCATCCACGTTAAACCGGCGCACCT--	5648		
Db	1730	uLysValAspGlyValAlaAlaTrpIleHisAlaSerHisValysAlaAlaGlyProTh	1750		
Qy	5649	-----CCGATTTCGGGTGGAAAGCCGA	5671		
Db	1750	rThrAsnGlnAspLeuSerAspSerProSerSerAspProSerArgTrpLysValG1	1770		
Qy	5672	AAAGACTGAAATCCCTTAAGCTTCGCTCCATCGC	5708		
Db	1770	mArgThrGlnAsnProLeuLysIleArgLeuSerArg	1782		
RESULT 7					
ID	AA05898				
XX	AA05898	standard; protein; 1784 AA.			
AC	AA05898;				
XX					
DT	24-OCT-2003	(revised)			
DT	27-AUG-2003	(revised)			
DT	25-MAR-2003	(revised)			
DT	29-NOV-1990	(first entry)			
XX					
DE	Gene product of first ORF of virus capable of inducing viraemia in feline				
DE	species encoded by clone 61E.				
XX					
KW	Viraemia; leukaemia; FeLV-A; clone 61C; clone 61E; ds.				
XX					
OS	Feline leukemia virus; strain A.				
FN	EP377842-A.				
XX					
PD	18-JUL-1990.				
XX					
PF	12-DEC-1989;	89EP-0012964.			
XX					
PR	13-DEC-1988;	88US-00284139.			
XX					
PA	(HARD ) UNIV HARVARD.				
PA	(COLS ) COLORADO STATE UNIV RES.				
PA	(HARD ) UNIV HARVARD.				
XX					
PI	Hoover EA, Mullins JI;				
XX					
DR	WPI; 1990-218326/29.				
DR	N-PSDB; AAQ05252.				
XX					
PT	DNA encoding leukaemia virus-A subtype - producing AIDS type disease in				
PT	cats, used to test drugs and vaccines.				
XX					
PS	Claim 6; Fig 1.1-1.7; 23pp; English.				
XX					
CC	Variant of viral genome induces immunodeficiency in cats similar to AIDS.				
CC	May be used in research, especially in testing drugs and vaccines against				
CC	viraemia and feline leukaemia viruses. (Updated on 25-MAR-2003 to				
CC	correct PA field.) (Updated on 27-AUG-2003 to correct OS field.) (Updated				
CC	on 24-OCT-2003 to standardise OS field)				
XX					
SQ	Sequence 1784 AA;				
Alignment Scores:					
Pred. No.:	0	Length:	1784		
Score:	5229.00	Matches:	1028		
Percent Similarity:	70.3%	Conservative:	235		
Best Local Similarity:	57.2%	Mismatches:	410		
Query Match:	35.7%	Indels:	123		

DB:	2	Gaps:	29
US-10-723-552-3 (1-8132) x AAR05898 (1-1784)			
QY	512	TCACGTGTCGTGATCTGTGTGGTTTCGTGTTTGTGTGTCTTTGTCTGTGTGTGTCCTGTGT	571
Db	47	SerArgAlaAlaArgLeuValIlePheCysLeuValAlaSerPheLeuVal---ProCys	65
QY	572	CTACAGTTTT-----AATATGGGACAGCGGTGACGACCCCTCTTAGTTT	616
Db	66	Leu-ThrPheLeuIleAlaGluThrValMetGlyGlnThrIleThrThrProLeuSerLe	85
QY	617	GACTCTCGACATTTGGACTGAAGTTAAATCCAGGGCTCATATAATTGTGACTTCAGGTTAA	676
Db	85	uThrLeuAephIstRpSerGluValAArgAlaAArgAlaAHisAenGlnGlyValGluValAr	105
QY	677	GAGGGACCTTGGCAGACTTCTGTGTCTCTGAATGCCGACATTCGATCTTGATGGCC	736
Db	105	gLYsLYsYstRpIleThrLeuCysGluAlaGluTrpValMetMetAsnValGlyTrpPr	125
QY	737	ATCAGAGGGACCTTTAAATCTGAGATATCTCGCTGTTAAAGCAGTATTTTTTCAGAC	796
Db	125	oArgGluGlyThrPheSerLeuAspAnIleSerGlnValGluLysLysIlePheAlaPr	145
QY	797	TGGACCCGGCTCTCATCCGATCAGGAGCCCTATATCTTACGTGGCAAGATTGGCAGA	856
Db	145	oGlyProTyrgLYsIstProAspGlnValProTyrgIleThrThrTrpArgSerLeuAlaTh	165
QY	857	GGATCTCCGCCATGGGTAAACCATCGCTGAATAAGCAAGAAGACCGAGTCCCCGAAT	916
Db	165	rAspProSerTrpValArgProPheLeuProProLysThrProThrPro-----	183
QY	917	TCTGGCTCTTGGAGAGAAAAACAACACTCGGCTGAAAAAGTCAAGCCCTCTCCT	971
Db	184	-----LeuProGlnProLeuSerProGlnProSerAl	194
QY	972	-----CATATCTACCCCGAGATT-----GAGGAGCCACCGCC	1003
Db	194	aProLeuThrSerSerLeuTygProValLeuProLysSerAspProProLysProProVa	214
QY	1004	TTGGCCCGAACCCTCACTGTCTCC-----CGACCCCTTAA	1039
Db	214	lLeuProProAspProSerProSerProLeuIleAspLeuLeuThrGluGluProProTy	234
QY	1040	TCTGGCACAGGTCGCGGAGGACCTTTGCCCTCTCGAGCTCCG-----GCGGT	1093
Db	234	r-----ProGlyGlyHisGlyPro-----ProProSerGlyProArgThrProTh	249
QY	1094	GGAGGACCTCTCGCAGGAGACTCGGAGCCGGAGGGCGCCACCCCGGAGCGGACAGACA	1153
Db	249	rAlaserProIleAlaSerArgLeuArgGluArgArgGluAsnPro-----AlaGlu	267
QY	1154	GATCGGCACATTACCGTGGCAGCTAGGGCCCTCCACACCGGGGGGCCAATTGACGC	1213
Db	267	uSerGlnAlaLeuProLeuArg-----GluGlyProAsnAsnArgPr	281
QY	1214	CCTCCAGTATTGGCCCTTTCTCTCTGCAGATCTCTATAATTGGAAAACTAACATCCCC	1273
Db	281	o-----GlnTyrrTrpProPheSerAlaSerAspLeuTygAsnTrpLysSerHisAsnProPr	300
QY	1274	TTTTCTCGAGGATCCCCAACCCCTCACGGGGTGGTGGAGTCCCTTATGTTCTCTCACCA	1333
Db	300	oPheSerGlnAspProValAlaLeuThrAsnLeuIleGluSerIleLeuValThrHisG	320
QY	1334	GCCTACTTGGGATGATTGTCAACAGCTGTGTCAGACACTCTTTCACAACCGAGGAGCGAGA	1393
Db	320	nProThrTrpAspAspCysGlnGlnLeuLeuGlnAlaLeuLeuThrGlyGluGluArgG	340
QY	1394	GAGAACTTATTAGAGCTGAAAAAATGTTCTTGGGGCCGACGGGCGACCCACCGCGTT	1453
Db	340	nArgValLeuLeuGluAlaArgLysGlnValProGlyGluAspGlyArgProThrGlnLe	360
QY	1454	GCAAAATGAGATTGACATGGGATTTCCTCTAACTCGCCCGCGGTGGGACTACAAACGGC	1513

Db	360	uProAsnValIrlEAspGluLaPheProLeuThrArgProAsnTrpAspPheArgThrPr	380
QY	1514	TGAAGGTAGGAGAGCTTGAATACTATCCAGGCTCTGGTCGGGTCTCCGGGGCGC	1573
Db	380	oAlaGlyArgGluHlsLeuArgLeuThrArgGlnLeuLeuLeuAlaGlyLeuArgGlyAl	400
QY	1574	CTCAAGACGGCCCACTAATTTGGCTAAGGTAAAGAGATGATCAGGAGCCGAATGAACC	1633
Db	400	aAlaArgArgProThrAsnLeuAlaGlnValLysGlnValGlnGlyLysGluGluTh	420
QY	1634	CCCTCTGTTTTTCTTGAGAGGCTCTTGGAAAGCCTTCAGCGGTATACCCCTTTTGNATCC	1693
Db	420	rProAlaSerPheLeuGluArgLeuLysGluAlaTyrArgMetTyrThrProTyrAspPr	440
QY	1694	CACCTCAGAGGCCCAAAAGCCCTCAGTGGCTTTGGCCCTTTATAGGACAGTCAGCCTTGGGA	1753
Db	440	oGluAspProGlyGlnAlaIaSerValIleLeuSerPheIleTyrGlnSerSerProAs	460
QY	1754	TATTAGAAAGAAGCTTCAGAGACTGGAAAGGTTTACAGGAGGCTCAGTTACGTGATCTAGT	1813
Db	460	pIleArgAsnLysLeuGlnArgLeuGluGlyLeuGlnGlyPheThrLeuSerAspLeuLe	480
QY	1814	GAAGGAGCGAGAGAAAGTATATTACAAAAGGAGACAGAAAGAGAAAGGAAACAAGAAA	1873
Db	480	uLysGluAlaGluLysIleTyrAsnLysArgGluThrProGluGluArgGluGluArg--	499
QY	1874	AGAGAGAGAAGAGAGCAAGAGGAGGAAAGACGCTAATAAAGCGACAGAGAAGATTGAC	1933
Db	500	-----LeuTrpGlnArgGlnGluGluArgAspLysArgHisLysGluMetTh	516
QY	1934	TAAGATCTTGCTGCAGTGGTTGAAGGAAAAGCAATACGGAAGAGAGAGA-----	1985
Db	516	rLysValLeuAlaThrValValAlaGlnAsnArgAspLysAspArgGluGluSerLysLe	536
QY	1986	----GATTTTAGGAAATTAGTTCAGGCCCTAGACAGTCAGGGAACCTGGGCAATAGGAC	2041
Db	536	uGlyAspGlnArgLysIle-----	542
QY	2042	CCCACTCGACAGAGCAACATGTGCATATTCTGAAGAAGAGGACACTGGCAAGGACTG	2101
Db	543	-ProLeuGlyLysAspGlnCysAlaTyrCysLysGluLysGlyHisIleTrpValArgAspC	562
QY	2102	CCCCAAGAGGGAACAAAGAGCA-----AGGATCTAGCTCTAGAGAAGATAA	2152
Db	562	sProLysArgProArgLysLysProAlaAsnSerThrLeuLeuAsnLeuGluAsp***G	582
QY	2153	AGATTAGGGAGACGGGTTCGAGCCCTCCCGAGCCACGAGGTAACTTTGAAGGTGGA	2212
Db	582	user-----GlnGlyGlnAspProProGluProArgIleThrLeuLysIleG	599
QY	2213	GGGCAACCACTGAGTTCCTGGTTGATACCGGAGCGAAACATTTCAGTGTCTACTACAGCC	2272
Db	599	yGlyGlnProValThrPheLeuValAspThrGlyAlaGlnHiserValLeuThrArgPr	619
QY	2273	ATTAGGAAACATAAAGATAAAATCCTGGGTGATGGGTGCCACAGGCAACAAACAGTA	2332
Db	619	oAspGlyProLeuSerAspArgThrAlaLeuValGlnGlyAlaThrGlySerLysAsnTy	639
QY	2333	TCCATGACTACCCGAAGAACAGTTGACTTTGGGAGTGGGACGGGTAAACCCTCGTTTCT	2392
Db	639	rArgTrpThrThrAspArgValGlnLeuAlaThrGlyLysValThrHiserPheLe	659
QY	2393	GGTCATACCTGAGTGGCCAGCACCCCTCTTAGGTAGAGACTTATTCACCAAGATGGGAGC	2452
Db	659	uTyrValProGluCysProTyrProLeuLeuGlyArgAspLeuLeuThrLysLeuLysAl	679
QY	2453	ACAAATTTCTTTTGA-- --CAAGGGAACACAGAAGTGTCTGCAAATAACAAACCTATCAC	2509
Db	679	aglnIleHisPheThrGlyGluGlyAlaAsnValValGlyProArgGlyLeuProLeuG	699
QY	2510	TGTTGTACCCCTCCAATTAGATACGGAATATCGACTATATCTCTCCCTAGTAAAGCCCTGA	2569

Db 699 nValLeuThrLeuGlnLeuGluGluGluTyArgLeuPheGluProGluSerThrGlnIly 719  
QY 2570 TCAAAATATACAAATCTGGTGTGAACAGTTCCTCCCAAGCCTGGGCGAGAAACCCGACGGAT 2629  
Db 719 sGlnGluMetAspIleTrpLeuLysAsnProGlnAlaTrpAlaGluThrGlyGlyMe 739  
QY 2630 GGGTTTGGCAACCAAGTTCCTCCCAACAGTATTCACTGAAGCCAGTGGCCACACAGT 2689  
Db 739 tGlyMetAlaHisCysGlnAlaProValLeuGlnLeuLysAlaThrAlaThrProI 759  
QY 2690 GTCAAGTCAGACAGTACCTCTGAGTAAAGAGCTCAAGAGGAATTCGGCCGATCTCCA 2749  
Db 759 eSerIleArgGlnTyProMetProHisGluAlaTyArgGlnGlyLeuProHisIleAr 779  
QY 2750 AGATTAAATCAACAGCGCATCTAGTTCCTGTCCTCAATCTCCCTGGAATACTCCCTGCT 2809  
Db 779 gArgMetLeuAspGlnGlyLeuLysProCysGlnSerProTrpAsnThrProLeuLe 799  
QY 2810 ACCGGTTAGAAAGCCTGGGACTAATGACTATCCACAGTACAGGACTTGAGAGAGTCAA 2869  
Db 799 uProValLysProGlyThrLysAspTyArgProValGlnAspLeuArgGluValAs 819  
QY 2870 TAAACGGGTGCAGATATACACCAACAGTCCCGAACCTTATAACCTCTGTGTGCTCT 2929  
Db 819 nLysArgValGluAspIleHisProThrValProAsnProTyAsnLeuLeuSerThrLe 839  
QY 2930 CCCACCCCAACGAGCTGGTATACAGTATTGGACTTAAAGGATGCCTTCTTCGCTGAG 2989  
Db 839 uProProSerHisProTrpTyThrValLeuAspLeuLysAspAlaPheCysLeuAr 859  
QY 2990 ATTACACCCCACTAGCAACCACTTTTTCCTTCGATTCGAATGGAGATCCAGGTACGGGAAG 3049  
Db 859 gLeuHisSerGluSerLeuLeuPheAlaPheGluTrpArgAspProGluIleGlyLe 879  
QY 3050 AACCGGGAGCTCACCTGACCGCTGCTCCCAAGGGTTCAGAACTCCCGACCATCTT 3109  
Db 879 uSerGlyGlnLeuThrTrpArgLeuProGlnGlyPheLysAsnSerProThrLeuPh 899  
QY 3110 TGACGAAGCCCTACACAGAGACCTGGCCCACTTCAGATCCCAACACCTTCAGGTGACCCT 3169  
Db 899 eAspGluAlaLeuHisSerAspLeuAlaAspPheArgValArgTyProAlaLeuValle 919  
QY 3170 CTTCCAGTACGTGGATGACTGCTTCTGGCGGAGCCCAACAAAGGACTGCTTGAAGG 3229  
Db 919 uLeuGlnTyValAspLeuLeuAlaAlaThrArgThrGluCysLeuGluG 939  
QY 3230 CACGAAGGCACCTACTGCTGAATTCCTGACCTAGCTACAGGCTCTGCTAAGAGGC 3289  
Db 939 yThrLysAlaLeuLeuGluThrLeuGlyAsnLysGlyTyArgAlaSerAlaLysAl 959  
QY 3290 CCAGATTTTCAGGAGAGAGTAACTACTTGGGGTACAGTTTGGCGGAGCGGCGAGATG 3349  
Db 959 aGlnIleCysLeuGlnGluValThrTyLeuGlyTySerLeuLysAspGlyGlnArgTr 979  
QY 3350 GCTGACGGAGGACGAGAAACCTGTAGTCCAGATACCGGCCCCCAACACAGCCAAACA 3409  
Db 979 pLeuThrLysAlaArgLysGluAlaIleLeuSerIleProValProLysAsnProArgI 999  
QY 3410 RATGAGAGATTTTGGGACAGCTGATTTTCAGACTGCTGGATCCCGGGTTCGAC 3469  
Db 999 nValArgGluPheLeuGlyThrAlaGlyTyCysArgLeuTrpIleProGlyPheAlaG 1019  
QY 3470 CTTAGCAGCCCACTCTACCCGCTAACCAAGAAAAAGGGGAATTCCTCGGGCTCCTGA 3529  
Db 1019 uLeuAlaAlaProLeuTyProLeuThrArgProGlyThrLeuPheGlnTrpGlyThrG 1039  
QY 3530 GCACAGAGGATTTGATGCTATCAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCT 3589  
Db 1039 uGlnGlnLeuAlaPheGluAsnIleArgLysAlaLeuLeuSerProAlaLeuGlyLe 1059  
QY 3590 CCTCAGCTAACTAACCTTTACCTTTATGTGATGAGCTAAGGAGTAGCCCGGG 3649  
Db 1059 uProAspIleThrLysProPheGluLeuPheIleAspGluAsnSerGlyPheAlaLysG 1079

QY 3650 AGTTTTAACCCAAACCTAGGACCATGAGAGACCTGTGCGCTACTCTGTCAAAGAGCT 3709  
Db 1079 yValLeuValGlnLeuLysLeuGlyProTrpLysArgProValAlaTyLeuSerLysLysLe 1099  
QY 3710 CGATCTGTAGCAGTGGTGGCCCATATGCTTGAAGGCTATCGAGCTGTGTGCCATACT 3769  
Db 1099 uAspThrValAlaSerGlyTrpProCysLeuArgMetValAlaAlaIleAlaIleLe 1119  
QY 3770 GGTCAAGGACGCTGACAAATTCAGCTTTGGACAGATATAACTGTAAATAGCCCCCATGC 3829  
Db 1119 uValLysAspAlaGlyLysLeuThrLeuGlyGlnProLeuThrIleLeuThrSerHisPr 1139  
QY 3830 ATTGGAGAACTCGTTTCGGCAGCCCCCAGACCGATGGATGATCCCAAGCCCGCATGACCCA 3889  
Db 1139 oValGluAlaLeuValArgGlnProProAsnLysTrpLeuSerAsnAlaArgMetThrHi 1159  
QY 3890 CTATCAAGCCCTGCTTCTC---ACAGAGAGGTTCAGTTCGGTCCACACGCGCTCTCAA 3946  
Db 1159 sTyrglnAlaMetLeuLeuAspAlaGluArgValHisPheGlyProThrValSerLeuAs 1179  
QY 3947 CCTGCGCACTCTCTGCTGGAAGAGACTGATGAACCA-----GTGACTCATGATGCCA 4000  
Db 1179 nProAlaThrLeuLeuProLeuProSerGlyLysProProArgLeuSerProAsp----- 1197  
QY 4001 TCAACTATTGATTGAGGAGACTGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGG 4060  
Db 1198 -----LeuAlaGluThrMetAlaGlnThrAspLeuThrAspGlnProLeuProAs 1214  
QY 4061 AGAAGTGTCACTGCTGTTCACTGACGGAAGCAGTATATGGTGGAGAGTGAAGGATGCC 4120  
Db 1214 pAlaAspLeuThrTrpTyThrAspGlySerSerPheIleArgAsnGlyGluArgLysAl 1234  
QY 4121 TGGGGCGGGTGGCGGACCGCAGCATCTGGGCGCAGCAGCTCCCGGAAGGAAC 4180  
Db 1234 aGlyAlaAlaValThrThrGluSerGluValIleTrpAlaAlaSerLeuProGlyTh 1254  
QY 4181 TTCAGCAAAAAGGCTGAGCTCATGCGCCTCACGCAAGCTTTTGGCGCTCCGCGAAGGAA 4240  
Db 1254 rSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaLysGlyLy 1274  
QY 4241 ATCCATAAATTTATAGCAGCAGAGTATGCTTTTGGAGTGCACACGTACATGGGC 4300  
Db 1274 sLysLeuThrValTyThrAspSerArgTyArgAlaPheAlaThrAlaHisValHisGlyG 1294  
QY 4301 CATCTATAAACAAAGGGGTGCTTACTCTCAGCAGGAGGGAATAAAGAAACAAGAGGA 4360  
Db 1294 uIleTyArgArgArgGlyLeuLeuThrSerGluGlyLysGluIleLysAsnLysAsnG 1314  
QY 4361 AATTTCTAAGCCTATTAGAACCGCTACATTTACAAAAGGCTAGCTATTATACACTGTC 4420  
Db 1314 uIleLeuAlaLeuLeuGluAlaLeuPheLeuProLysArgLeuSerIleIleHisCysPr 1334  
QY 4421 TGGCATCAGAAAGCTAAAGATCTCATATCCAGAGAAACACAGATGGCTGACGGGTGC 4480  
Db 1334 oGlyHisGlnLysGlyAspSerProGlnAlaLysGlyAsnArgLeuAlaAspThrAl 1354  
QY 4481 CAAGCAGGCACCC-----CAGGGTGTAACTTCCTGCTCT-----ATAATAGA 4522  
Db 1354 aLysLysAlaAlaThrGluThrGlnSerSerLeuThrIleLeuProThrGluLeuIleG 1374  
QY 4523 AATGCCAAAGCCCGACAGACCCAGCAGCAGTACACCTTAGAAGACTGGCAAGAGATAA 4582  
Db 1374 uGlyProLys-----ArgProProTrpGluTyArgAspSerAspLeuAspLeuValG 1392  
QY 4583 AAAGATAGACCATTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGA 4642  
Db 1392 nLysLeuGluAlaHisTyGluProLysArgGlyThr---TrpGluTyArgGlyLysTh 1411  
QY 4643 AATCTGCCCCCAAGAAAGGGTGAATATGTCCAACAGATACATCGTCTTAACCCACT 4702  
Db 1411 rIleMetProGluLysTyArgAlaLysGluLeuIleSerHisLeuHisLysLeuThrHisLe 1431

```
QY 4703 AGGAACCTAAACACCTGCAGCAGTTGGTCAGAACATCC-----CCTTA 4744
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1431 userAlaArgLysMetLysThrLeuLeuGluArgGluThrGlyPheTyrlLeuProAs 1451
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 4745 TCATGTTCTGAGGCTACAGGATGGCTGACTCGGTGGTCAAAACATTGTGCGCCGCA 4804
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1451 nArgAspLeuHisLeuArgGlnValThrGluSer-----CysArgAlaCysAl 1467
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 4805 GCTGGTTAATGCTTAATCCTTCCAGATGCTCCAGGGAAGAGACTAAGGGAGACCCACC 4864
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1467 aGlnIleAsnAlaGlyLysIleLysPheGlyLeuAspValArgAlaArgGlyArgPr 1487
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 4865 AGCGCTCACTGGGAAGTGCATCTCACTGAGTAAAGCCGCTAAATACGGAACAATA 4924
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1487 oGlyThrHisTrpGluValAspPheThrGluIleLysProGlyMetTyrlTyrlsTy 1507
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 4925 CCTATTGGTTTGTAGACACCTTTTTCAGGATCGGTAGAGCTTATCCTACTAAGAAGA 4984
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1507 rLeuLeuValPheIleAspThrPheSerGlyTrpAlaGluAlaTy:ProAlaLysHisG 1527
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 4985 GACTTCAACCGTGGTGTAAATAAATACTGGAAGAAATTTTCCAGATTTCGATACC 5044
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1527 uThrAlaLysValAlaLysLysLeuLeuGluIlePheProArgTyrlGlyIlePr 1547
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5045 TAAGGTAATAGGCTCAGACATCGTCCAGCTTTTGTTCGCCAGTAAAGTCAGGACTGGC 5104
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1547 oGlnValLeuGlySerAspAsnGlyProAlaPheIleSerGlnValSerGlnSerValAl 1567
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5105 CAAGATATTGGGATGATTGGAACTGCATTTGTGCATACAGACCCCAAGCTCAGGACA 5164
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1567 aThrLeuLeuGlyIleAsnTrpLysLeuHisCysAlaTyArgProGlnSerSerGlyG 1587
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5165 GCTAGAGAGATGAATAGACACCTTAAGAGACCTTACTTAATACCCCGAGACTGG 5224
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1587 nValGluArgMetAsnArgSerIleLysGluThrLeuThrLysLeuLeuGluThrG 1607
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5225 CGTTAATGATTGATAGCTCTCGCTGCTTTGTGCTTTTGTAGGGTTAGGAACCCCTGG 5284
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1607 ySerLysAspTrpValLeuLeuLeuProLeuValLeuTyArgValArgAsnThrProG 1627
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5285 ACAGTTTGGGCTGACCCCTATGAATTAATCTACGGGGGACCCGCCATTTGGTAGAAAT 5344
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1627 yProHisGlyLeuThrProPheGluIleLeuTyArgGlyAlaProProMetAlaHisPh 1647
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5345 TGCCTTCTGATACATAGTGCACGTGCTGCTTCCAGACCTTTCTTCTAGGCTCAGGC 5404
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1647 ePheAspThrAspIleSerThrPheAlaThrSerProThrMetGlnAlaHisLeuArgAl 1667
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5405 ACTTGAGTGGTGAGACACGACGCTGGAGGCAACTCCGGAGGCGCTACTCAGGAGGAGG 5464
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1667 aLeuGlnLeuValGlnGluGluGluGlnArgProLeuAlaAlaAlaTyArgGluLysLe 1687
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5465 AGACTTGCAG---ATCCCCACATCGTTTCCAAAGTGGAGATTACGTCTAGCTTAGACGCCA 5521
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1687 uGluThrProValValProHisProHisProLysProGlyAspSerValTrpValArgArgHi 1707
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5522 CCGTCGAGAAACCTCGAGACTCGGTGGAAGGCCCTTATCTCGTACTTTTGACACACC 5581
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1707 sGlnThrLysAsnLeuGluProArgTrpLysGlyProHisIleValLeuLeuThrThrPr 1727
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5582 AACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGATCCATGATCCACGCTTAAACCGGC 5641
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1727 oThrAlaLeuLysValAspGlyValAlaAlaTrpIleHisAlaSerHisValLysAlaAl 1747
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5642 GCCACCT-----CCCCGCTCGGGGTG 5662
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1747 aGlyProThrThrAsnGlnAspLeuSerAspSerProSerSerAspAspProSerArgTr 1767
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 5663 GAAAGCGGAAAGACTCAAAATCCCTTAAGCTTCGCTCCATCCG 5708
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1767 pLysValGlnArgThrGlnAsnProLeuLysIleArgLeuSerArg 1782
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 8
```

```
AAB73284
ID AAB73284 standard; protein; 1069 AA.
AC AAB73284;
XX
XX
XX 23-MAY-2001 (first entry)
XX
XX Defective retroviral genome protein #3 isolated from PK-15 cell line.
XX
XX Retrovirus; graft transplantation; xenotransplantation; PK-15 cell line.
XX
XX Unidentified.
XX
XX US6190861-B1.
XX
XX 20-FEB-2001.
XX
XX 13-DEC-1996; 96US-00766528.
XX
XX 14-DEC-1995; 95US-00572645.
XX (GEHO ) GEN HOSPITAL CORP.
XX Fishman JA;
XX
XX WPI; 2001-256211/26.
XX N-PSDB; AAF77726.
XX
XX Assessing risk of endogenous retroviruses in clinical practice and in
XX xenotransplantation, comprises using probe sequences derived from swine
XX or miniature swine retroviral genome.
XX
XX Disclosure; Fig 2; 127pp; English.
XX
XX The present invention relates to a method for screening a cell or tissue
XX for the presence or expression of a retrovirus (RV), comprising
XX contacting a target nucleic acid from the cell or tissue with a second
XX nucleic acid from the present invention (e.g. AAF77726 or a fragment
XX thereof). The method is useful for RV detection and to assess graft
XX transplantation risk. Screening of animals allows the elimination of
XX donors with active replication of known viruses. Inactive proviruses can
XX be detected and inactivated, allowing identification and elimination of
XX potential human pathogens derived from swine in a manner not possible in
XX the outbred human organ donor population and is important to the
XX development of human xenotransplantation
XX
XX Sequence 1069 AA;
SQ
Alignment Scores:
Pred. No.: 0 Length: 1069
Score: 4889.00 Matches: 930
Percent Similarity: 89.8% Conservative: 46
Best Local Similarity: 85.6% Mismatches: 62
Query Match: 33.4% Indels: 49
DB: 4 Gaps: 8
US-10-723-552-3 (1-8132) x AAB73284 (1-1069)
QY 4311 CAAAGGGGGTGTCTTACCTCAGCAGGAGGAAATAAGAACAGGAAATCTTAAGC 4370
Db 1 LysArgGlyLeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluLeuSer 20
QY 4371 CTATTAGAGCGGTACATTTACCAAAAGCGTAGCTATTATACACTGTCTCGACATCAG 4430
Db 21 LeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIleHisCysProGlyHisGln 40
QY 4431 AAAGCTAAAGATCTCATATCCAGAGAAACACAGATGGCTGACCGGTTGCCAAGCAGGCA 4490
Db 41 LysAlaLysAspLeuIleSerArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAla 60
QY 4491 GCCAGGGGTGTTAAACCTTCTGCTTATATAGAAATGCCCAAGCCCGACAGACGCA 4550
Db 61 AlaGlnAlaValAsnLeuLeuProIleIleGluThrProLysAlaProGluProArg 80
```

QY 4551 CAGTACACCTAGAGACTGGCAAGATATAAAGATAGACCAAGTTCTCTGAGACTCCG 4610  
 DB 81 GlnTyr-ThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrPro 100  
 QY 4611 GAAGGACCTGCTATACCTCAGATGGGAAGAAATCTGCCCCACAAAGAGGTTAGAA 4670  
 DB 101 GluGlyThrCysTyrThrSerTyrGlyLysGluIleLeuProHisLysGluGlyLeuGlu 120  
 QY 4671 TATGTCACACAGATACATCGTCTAAACCCACCTAGGAACATAAACCTGCAGCAGTTGGTC 4730  
 DB 121 TyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnLeuVal 140  
 QY 4731 AGAACATCCCTTATCATGTTCTGAGCTACACGAGTGGCTGACTCGTGGTCAACAT 4790  
 DB 141 ArgThr-SerProTyrHisValLeuArgLeuProGlyValAlaAspSerValValLysHis 160  
 QY 4791 TGTGTCCTCCGCTGCTGTTTAAATGCTTAATCTCCAGATGCTCCAGGAAGACTA 4850  
 DB 161 CysValProCysGlnLeuValAsnAlaAsnProSerArgIleProGlyLysArgLeu 180  
 QY 4851 AGGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTCAGGTAAAGCCGGCTAAA 4910  
 DB 181 ArgGlySerHisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLys 200  
 QY 4911 TAGGAAACAAATACCTATTGTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTAT 4970  
 DB 201 TyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaTyr 220  
 QY 4971 CCTACTAAGAAAGACTCAACCGTGGTGGCTTAAATAAATACTGAAGAAATTTTCCA 5030  
 DB 221 ProThrLysLysGluThrSerThrValValAlaLysLysIleLeuGluIlePhePro 240  
 QY 5031 AGATTTGGAATACCTAAGGTAAATAGGTGACAGCAATGGTCCAGCTTTGTGCCAGGTA 5090  
 DB 241 ArgPheGlyIleProLysValIleGlySerAspAsnGlyProAlaPheValAlaGlnVal 260  
 QY 5091 AGTCAGGAGTGGCCAAAGATATTGGGATTTGATGGAAACTGCAATTTGCATACAGACC 5150  
 DB 261 SerGlnGlyLeuAlaLysIleLeuGlyIleAsp\*\*LysLeuHisCysAlaTyrArgPro 280  
 QY 5151 CAAGCTCAGGACAGTATAGAGGATGAATAGAACCATTAAGAGACCTTACTAATATG 5210  
 DB 281 GlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeu 300  
 QY 5211 ACCGCGAGACTGGCGTTAATGATGATAGCTCTCCTGCTGCTGCTGCTTTTGGGTT 5270  
 DB 301 ThrThrGluThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeuPheArgVal 320  
 QY 5271 AGGAACACCCCTGGACAGTTTGGGCTGACCCCTATGAATTTACTTACGGGGGACCCCTC 5330  
 DB 321 ArgAsnThrProGlyGlnPheGlyLeuThrProTyrLysLeuLeuTyrGlyProPro 340  
 QY 5331 CCAATGGTGAATTCCTCTCTGTAATAGTGTGAGTGTGCTGCTTTCCAGCTTTGTTTC 5390  
 DB 341 ProLeuAlaGluIleAlaPheAlaHisSerAlaAspValLeuLeuSerGlnProLeuPhe 360  
 QY 5391 TCTAGGCTCAAGCACTTCAGTGGGTGACACACGAGCTGGAGGCAACTCCGGGAGGCC 5450  
 DB 361 SerArgLeuLysAlaLeuLeuThrProValArgGlnArgAlaTrpLysGlnLeuArgGluAla 380  
 QY 5451 TACTCAGGAGGAGGAGACTTGAGATCCCACATCGTTTCCAAAGTGGGAGATTCAGTCTAC 5510  
 DB 381 TyrSer--GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTyr 399  
 QY 5511 GTTAGCGCCACCTGCAGAAACCTCGAGACTCGGTGGAAGGCCCTTATCTGACTT 5570  
 DB 400 ValArgArgHisArgAlaGlyAsnLeuGluThrArg\*\*LysGlyProTyrLeuValLeu 419  
 QY 5571 TTGACCACCAACGGCTGTGAAGTCGAAGGAATCTCCACCTGGATCCATCATCCAC 5630  
 DB 420 LeuThrThrProThrAlaValLysValGluGlyIle-Pro----- 432

QY 5631 GTTAAACCGCGCCACCTCCCGATTCCGGGTGAAAGCCGAAAGACTGAAATCCCTT 5690  
 DB 433 -----Le 433  
 QY 5691 AAGCTTCGCTCCATCGCTGGTTCCTTACTCTGTCAATAACTCTCAGACTAATGGTAT 5750  
 DB 433 userPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGlnValAsnGlyLys 453  
 QY 5751 GGCATATGAGAGCAGCTGAACCTCCCATAAACCTTATCTCTCACCTGGTAAATTAATCTGA 5810  
 DB 453 sArgLeuValAlaSerProAsnSerHisLysProLeuSerLeuThrTrpLeuLeuThrAs 473  
 QY 5811 CTCGGCACAGGTATTAAATATCAACACACTCAAGGGGAGGCTCTCTTATGAAACCTGGTG 5870  
 DB 473 pSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeuGlyThrTrpTr 493  
 QY 5871 GCCTGATCTATAGTTTGGCTCAGATCAGTTATTCCTAGTCTG-----ACCTCACC 5921  
 DB 493 pProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAspGlnAlaThrPr 513  
 QY 5922 CCAGATATCTCCATGCTCACGATTTCCTGCTTTCGCCAGGACCAACCAATATGAAA 5981  
 DB 513 oProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProProAsnAsnGluGl 533  
 QY 5982 ACATTCGGAAATCCAGAGATTCTTTTGTAAACAATGGAACCTGTAACTCTAATGA 6041  
 DB 533 uTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysIleThrSerAsnAs 553  
 QY 6042 TGGATATTGGAATGGCCAACTCTCAGCAGATAGGGTAAGTTTTCCTTATGTCAACAC 6101  
 DB 553 pGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSerPheValAsnAs 573  
 QY 6102 CTATACCACTCTGGACAATTTAATTAC-----CTGACCTGG----- 6138  
 DB 573 nProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAspTrpGlnGlnAr 593  
 QY 6139 -----ATTAGAACTGGAGCCCAAGTCTCTCTCTCAGACCTAGATTACCT 6185  
 DB 593 gValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAspLeuAspTyrLe 613  
 QY 6186 AAAATAAGTTTCACTGAGAAAGGAAACAGAAATATCTTAAATGGGTAAATGGTAT 6245  
 DB 613 uLysIleSerPheThrGluLysGlyLysGlnGlnAsnIleGlnLysTrpValAsnGlyL 633  
 QY 6246 GTCTGGGGAATGGTATATTATGGAGGCTCGGTAAACAACACAGGCTCCATCTTAATCT 6305  
 DB 633 eSerTrpGlyIleValTyrTyrGlySerGlyArgLysLysGlySerValLeuThrIle 653  
 QY 6306 TCGCTCTCAAATA-----AACCACTGGAGCTCCAATGGCTATAGGACCAATAACGCTCT 6362  
 DB 653 eArgLeuArgIleGluThrGlnMetGluProValAlaIleGlyProAsnLysGlyLe 673  
 QY 6363 GACGGGTCAAGACCCCAACCCAA-----GGACGAGGACCATCTCTCT---AA 6407  
 DB 673 uAlaGluGlnGlyProProIleGlnGluArgProSerProAsnProSerAspTyrAs 693  
 QY 6408 CATAACTTCTGATCAGACCCCACTGAGTCAACAGCAGCTAAATCGGGGCAAACT 6467  
 DB 693 nThrThrSerGlySerValProThrGluProAsnIleThrIleLysThrGlyAlaLysLe 713  
 QY 6468 TTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCAGACTCCAGAGGCTACCTC 6527  
 DB 713 uPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThrSe 733  
 QY 6528 TTCTTTGGCTATCGTTAGCTTCGGGCCCACTTACTATGAGGATCGGTAGAGAGG 6587  
 DB 733 rSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMetAlaArgGlyGl 753  
 QY 6588 GAAATTCATGTGACAAAAGACATAGAGACCAATGCATCGGATCCCAATAAGCT 6647  
 DB 753 yLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGlnAsnLysLe 773  
 QY 6648 TACCCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGAAAGTTTCCCCCATCCACCA 6707



```
Db      773 uThrLeuThrGluValSerGlyLysGlyThrCysIleGlyMetValProSerHisGI 793
QY      6708 ACACCTTTGTAAACACACTGAAGCCTTTAAATCAAACTCTGAGAGTCAATATCTGTGACC 6767
Db      793 nHisLeuCysAenHisThrGluAlaPheAenArgThrSerGluSerGlnTyrlLeuValPr 813
QY      6768 TGCTTATGACAGCTGGTGGGCATGTAAATACATGATTAACCCCTTGTTCCACCTTGGT 6827
Db      813 oGlylTyAspArgTrpAlaCysAenThrGlyLeuThrProCysValSerThrLeuVa 833
QY      6828 TTTTAACCAAACTAAAGATTTTGCATTATGTGTCCTCAAAATGTTCCTCCGAGTGTATTACTA 6887
Db      833 lPheAenGlnThrLysAspPheCysValMetValGlnIleValProArgValTyrlTy 853
QY      6888 TCCGAAAAGCAATCCTTGATGAATATGATCTACAGAAATCATCGACAAAGAGAGAAC 6947
Db      853 rProGluLysAlaValLeuAspGluTyrlAspTyrlArgTyrlAsnArgProLysArgGluPr 873
QY      6948 CATATCTCTGACACTTCTGCTGATGCTCGGACTTGGAGTGGCAGCAGGTGTAGAACAGG 7007
Db      873 oIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyValGlyThrGI 893
QY      7008 AACAGCTGCCCTGCTGACGGGACACAGCAGCTAGAAACAGGACTTAGTAACCTACATCG 7067
Db      893 yThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSerAsnLeuHisAr 913
QY      7068 AATTGTAACAGAGATCTCAAGCCTTAGAAAATCTGTGAGTAACCTGGAGGAATCCCT 7127
Db      913 gIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluGluSerLe 933
QY      7128 AACCTCTTATCTGAGTACTCTACAGATAGAGAGGTTAGATTATTATTCTTAA 7187
Db      933 uThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuLy 953
QY      7188 AGAAGAGAGATTATGTGACCTTGAAGAGGAATGCTGTTTATGTGATCATTCAGG 7247
Db      953 sGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrlValAspHisSerGI 973
QY      7248 GGCATCAGAGATCCCATGAAAGCTTAGAGAAAGTTGGAGAGCGTCGAAGGGGAA 7307
Db      973 yAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArgArgArgGluAr 993
QY      7308 GGAACCTACTCAAGGTGTTTCAGGATGCTTCACAGCTCTCTTGTGGCTACCCCT 7367
Db      993 gGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetThrThrLe 1013
QY      7368 ACTTTCCTGCTTAAACAGACCCCTTAATAGTCCTCCTCTCTACTCACAGTTGGGCCATG 7427
Db      1013 uLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThrValGlyProCy 1033
QY      7428 TATTATTAAACAGTTAATTCCTTCAATTAGAGAACGAATAGTCGATCCAGATCATGTT 7487
Db      1033 sLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnIleMetVa 1053
QY      7488 ACTTAGACACAGTACC 7506
Db      1053 lLeuArgGlnGlnTyrlGln 1059
```

## RESULT 9

```
AAW81572
ID  AAW81572 standard; protein; 1203 AA.
XX
AC  AAW81572;
XX
DT  01-MAR-1999 (first entry)
XX
DE  Mus dunni endogenous virus Pol protein.
XX
KW  MDEV; retrovirus; packaging cell line; gene transfer; gene therapy;
KW  vector; polymerase protein; Pol protein.
XX
OS  Mus dunni endogenous virus.
XX
```

```
XX
PN
XX
PD  12-NOV-1998.
XX
PF  08-MAY-1998; 98WO-US009452.
XX
PR  09-MAY-1997; 97US-0046140P.
PR  08-MAY-1998; 98US-00075272.
XX
PA  (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI  Miller AD, Wolgamot G, Bonham L;
DR  WPI; 1999-034718/03.
DR  N-PSDB; AAV69750.
XX
PT  New retroviral packaging cells - containing Mus dunni endogenous virus
PT  sequences to target cells, retrovirus gag and pol genes and a
PT  heterologous gene of interest.
XX
XX  Disclosure; Page 68-71; 85pp; English.
XX
CC  This is the amino acid sequence of the polymerase (Pol) protein of Mus
CC  dunni endogenous virus (MDEV), as deduced from the MDEV nucleotide
CC  sequence (see AAV69750). A cultured packaging cell is claimed which
CC  produces a replication-defective retroviral vector (RDRV) particle, where
CC  the packaging cell is a vertebrate cell capable of expressing and
CC  assembling retroviral proteins, comprising: (a) a first vector encoding a
CC  retroviral envelope protein having amino acid residues MDEV that direct
CC  binding of the retroviral particle to MDEV retroviral receptors on a
CC  target cell; and (b) a second vector encoding retrovirus gag and Pol
CC  proteins, where upon expression of the vectors in the packaging cell in
CC  the presence of a vector having a sequence of a heterologous gene of
CC  interest, a replication-defective retroviral particle is produced that
CC  binds to MDEV receptors of target cells. Also claimed are: (1) a cultured
CC  packaging cell for producing a RDRV particle; (2) methods for producing a
CC  RDRV particle comprising a heterologous gene of interest; (3) cultured
CC  packaging cell line PD223; and (4) a RDRV produced by a method as in (2).
CC  The MDEV receptor is present on a variety of cells rendering MDEV
CC  pseudotype packaging cells useful in methods of mammalian and
CC  particularly human gene transfer for gene therapy. The MDEV packaging
CC  cells are a stable and reproducible source of retroviral particles.
CC  Clones may be isolated from these populations that produce high titre
CC  virus. The packaging cell lines may be selected and cloned for other
CC  desirable properties, such as stability of in vivo growth, lack of
CC  production of helper virus, lack of reinfection by viral particles
CC  packaged in the cell, stability from genetic rearrangement and
CC  recombinational events, resistance to complement lysis, and improved
CC  ability to infect cells from higher mammals
XX
SQ  Sequence 1203 AA;
```

## Alignment Scores:

Pred. No.:	0	Length:	1203
Score:	4332.50	Matches:	822
Percent Similarity:	79.6%	Conservative:	129
Best Local Similarity:	68.8%	Mismatches:	222
Query Match:	29.6%	Indels:	21
DB:	2	Gaps:	7

US-10-723-552-3 (1-8132) x AAW81572 (1-1203)

```
QY      2160 GGGAGACGGGTTCGGACCCCTCCCGAGCCAGGTAACTTTGAAGTGGAGGGCAA 2219
Db      1 GlySerArgGlySerAspProLeuProGluProArgValThrLeuSerValGluGlyThr 20
QY      2220 CCAGTTGAGTTCCTGCTGTGATACCGGAGCGAATTCAGTGTCTACTACAGCCATTAGGA 2279
Db      21 ProValAsnPheLeuIleAspThrGlyAlaGluHisSerValLeuThrSerProLeuGly 40
QY      2280 AAACCTAAAAGATAAAAAATCCTCGGTGATGGTCCACAGGGCAACAAACATGATCCATGG 2339
```

Db 41 LysLeuGlySerLysArgThrIleValValGlyAlaThrGlySerLysLeuTyrProTrp 60  
QY 2340 ACTACCGAAGAACAGTGTGACTTTGGAGTGGGACGGGTAAACCCAGCTCGTTTCTGGTCATA 2399  
Db 61 ThrThrLysArgAlaLeuGlnIleAspLysAsnMetValThrHisSerPheLeuValIle 80  
QY 2400 CCTGAGTGGCCAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATGGGAGCACAAAT 2459  
Db 81 ProGluCysProAlaProLeuLeuGlyArgAspLeuLeuThrLysLeuLysAlaGlnVal 100  
QY 2460 TCTTTGAAACAGGGAAACAGAAAGTGTCTGCAATAACAAACCTACCTGCTGTTGACC 2519  
Db 101 GlnPheThrSerGluGlyProGlnValSerTrpGlyLysAlaProLeuAlaCysLeuVal 120  
QY 2520 CTCCAATTAGATGACGAATATCGACTATATCTCTCCCTAGTAAAGCCTGATCAAAATATA 2579  
Db 121 LeuSerThrGluGluGluTyrArgLeuHisGluGluGlnProLysGlyAlaAlaProLeu 140  
QY 2580 CAATTCGTGTTGAAACAGTTTCCCAAGCCTGGCGAGAAACCCAGGGATGGTTGGCA 2639  
Db 141 Asp---TrpValThrAlaPheProAsnValTrpAlaGluGlnAlaGlyMetGlyLeuAla 159  
QY 2640 AAGCAAGTTCCTCCCAACAGTTATTCAACTGAAGCCAGTGCCACACAGTGTCTCAGTCAGA 2699  
Db 160 LysGlnValProProValValValGluLeuLysAlaAspAlaThrProLysSerValArg 179  
QY 2700 CAGTACCCCTTGTAGTAAAGAGCTCAAGAGGAATTCGGCCGATGTCCAAAGATTAAATC 2759  
Db 180 GlnTyrProMetSerLysGluAlaLysGluGlyIleArgProHisIleArgLeuLeu 199  
QY 2760 CAACAGGGATCCTAGTCTCTGTCCTCAATCTCCCTCGGAATCTCCCTGCTACCGTTAGA 2819  
Db 200 AspGlnGlyLeuLeuValAlaCysGlnSerProTrpAsnThrProLeuLeuProValArg 219  
QY 2820 AAGCCTGGGACTAATGACTATCCAGCAGTACAGACTTGAGAGCTCAATAAACGGGTG 2879  
Db 220 LysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsnLysArgVal 239  
QY 2880 CAGGATATACCCCAACAGTCCCGAACCTTTAAGCTCTTGTCCTCTCCCAACCCCAA 2939  
Db 240 LeuAspIleHisProThrValProAsnProTyrAsnLeuLeuSerSerLeuProGlu 259  
QY 2940 CGGAGCTGGTATACAGTATTGACTTAAAGATGCTTCTCTGCTGAGATTACACCCC 2999  
Db 260 ArgThrTrpTyrThrValLeuAspLeuLysAspAlaPheCysLeuArgLeuHisPro 279  
QY 3000 ACTAGCCACCACTTTTGCCTTCGAATCGAGAGATCCAGGTACGGAGAACCCGGCAG 3059  
Db 280 LysSerGlnLeuLeuPheAlaPheGluTrpArgAspProGluGlyGlnThrGlyGln 299  
QY 3060 CTCACCTGGACCCGACTGCCCCAAGGTTCAAGAACTCCCGACCATCTTGACGAAGCC 3119  
Db 300 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAspGluAla 319  
QY 3120 CTACACAGAGACTGGCCAACTTCAGGATCCAAACCCCTCAGTGTACCCCTCTCCAGTAC 3179  
Db 320 LeuHisArgAspLeuAlaProPheArgAlaGlnAsnProGlnLeuThrLeuLeuGlnTyr 339  
QY 3180 GTGGATGACTGCTTCTGGGGAGGCCAACACAGGACTGTTAGAGCGACGAAGGCA 3239  
Db 340 ValAspAspLeuLeuIleAlaAlaSerLysGluLeuCysGlnGlnGlyThrGluArg 359  
QY 3240 CTACTGCTGGAATTGTCTGACTAGGCTACAGAGCTCTGCTAAGAGGCCAGATTGCT 3299  
Db 360 LeuLeuThrGluLeuGlyAsnLeuGlyTyrArgValSerAlaLysLysAlaGlnIleCys 379  
QY 3300 AGGAGAGAGGTAACTACTTTGGGGTACAGTTTGGGGACGGGCGGATGGCTGACGGAG 3359  
Db 380 GlnThrGluValIleTyrLeuGlyTyrThrLeuArgGlyGlyLysArgTrpLeuThrGlu 399  
QY 3360 GCACGGAAGAAACTGTACTCCAGATACCGGCCCCCAACACAGCCAAACAAATGAGAGAG 3419  
Db 400 AlaArgLysLysThrValMetIleProProThrProArgGlnValArgGlu 419

QY 3420 TTTTGGGACAGCTGGATTTTGCAGACTGTGATCCCGGGTTTTCGACCTTTAGCAGCC 3479  
Db 420 PheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAlaAla 439  
QY 3480 CACCTCTACCCGTAAACCAAGAAAGGGAATTTCTCTGGCTCTCTGAGCACCAAG 3539  
Db 440 ProLeuTyrProLeuThrArgGluGlyIleProPheGluTrpLysGluGluHisGlnArg 459  
QY 3540 GCATTTGATGCTATCAAAAGGCCCTGCTGAGCGCACCTGCTGCGCCCTCCCTGAGCTA 3599  
Db 460 AlaPheGluAlaIleLysSerSerLeuMetThrAlaProAlaLeuAlaLeuProAspLeu 479  
QY 3600 ACTAAACCCCTTACCTTTATGTGGATGAGCGTAAAGGAGTAGCCCGGGGAGTTTAAACC 3659  
Db 480 ThrLysSerPheValLeuTyrValAspGluArgAlaGlyIleAlaArgGlyValLeuThr 499  
QY 3660 CAAACCTTAGGACCATGGAGAAGACCTGTGCGCTACCTGTCAAAGAGCTCGATCTGTGA 3719  
Db 500 GlnAlaLeuGlyProTrpLysArgProValAlaTyrLeuSerLysLysLeuAspProVal 519  
QY 3720 GCCAGTGGTGGCCCATATGCTGAGGCTATCGAGCTGTGCGCATATCTGTCAAGGAC 3779  
Db 520 AlaSerGlyTrpProThrCysLeuLysAlaIleAlaValAlaLeuLeuIleLysAsp 539  
QY 3780 GCTGACAAATTGACTTTGGGACAGAAATAAATGTAATAGCCCCCATGATTGGAGAAC 3839  
Db 540 AlaAspLysLeuThrMetGlyGlnValThrValValAlaProHisAlaLeuGluSer 559  
QY 3840 ATCGTTGGCAGCCCGACAGCCGATGATGACCAAGCCCGCATGACCCACTATCAAGC 3899  
Db 560 IleValArgGlnProProAspArgTrpMetThrAsnAlaArgMetThrHisTyrGlnSer 579  
QY 3900 CTGCTTCTCACAGAGGGTCACTTCGCTCCACAGCGCTCTCAACCTGCCACTCTT 3959  
Db 580 LeuLeuLeuAsnAspArgValThrPheAlaProProAlaIleLeuAsnProAlaThrLeu 599  
QY 3960 CTGCTTGAAGAGACTGAAACGAGTCACTGATGTTGCTCACTCACTATTGATTGAGGAG 4019  
Db 600 LeuProLeuThrAsnAspSerValProValHisArgCysAlaAspIleLeuAlaGlu 619  
QY 4020 ACTGGGTCGCCAAGACCTTACAGACATACCGCTGACTGGAGAGTGTCTAACCTGGTTC 4079  
Db 620 IleGlyThrArgLysAspLeuThrAspGlnProTrpProGly---AlaProSerTyr 638  
QY 4080 ACTGACGGAAGCAGCTATGCTGTGNAAGTAGAGATGGCTGGGGCGGGTGGTGAC 4139  
Db 639 ThrAspGlySerSerPheLeuIleGluGlyLysArgAlaGlyAlaValValAsp 658  
QY 4140 GGGACCCCGCACCATCTGGGCCAGCAGCTGCGGAGGAACTTCAGCACAAAAGGCTGAG 4199  
Db 659 GlyLysLysValIleTrpAlaSerAlaLeuProGluGlyThrSerAlaGlnLysAlaGlu 678  
QY 4200 CTCATGGCCCTCACCAAGCTTTGCGCTGGCGAAGGAAATCCATAAACAATTTATACG 4259  
Db 679 LeuIleAlaLeuThrGlnAlaLeuArgGluAlaGluGlyLysIleIleAsnIleTyrThr 698  
QY 4260 GACACAGGTATGCTTTGCGACTCCACACGTACATGGGCCCATCTATATAACAAGGGGG 4319  
Db 699 AspSerArgTyrAlaPheAlaThrAlaHisIleHisGlyAlaIleTyrArgGlnArgGly 718  
QY 4320 TTGCTTACTCTAGCGAGGAGGAAATAAGAACAAAGAGGAATTTCTAGCTCTATTAGAA 4379  
Db 719 LeuLeuThrSerAlaGlyLysAspIleLysAsnLysGluGluIleLeuAlaLeuLeuGlu 738  
QY 4380 GCGGTACATTACCAAAAAGGCTAGCTATTATACACTGCTCTGGACATCAGAAAGCTAAA 4439  
Db 739 AlaIleHisAlaProLysLysValAlaIleIleHisCysProGlyHisGlnLysGlyGlu 758  
QY 4440 GATCTCATATCCAGAGGAACACAGATGCTGACCGGGTTGCCAAGCAGGCAGCCAGGT 4499  
Db 759 AspLeuValAlaLysGlyAsnArgMetAlaAspSerValAlaLysGlnValAlaGlnGly 778

```
QY 4500 GTTAACTCTCTGCTATATAGAAATGCCCC---AAAGCCCCAGAACCCAGACGACAGTAC 4556
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
779 AlaMetIleLeuThrGluLysGlyAsnProSerLysSerProGlu-----793
QY 4557 ACCCTAGAAGACTGGCAAGAGATATAAAAGATAGACAGTTCCTGAGACTCCGGAAGGG 4616
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
794 -----AapGluAenTyrrAspIleLysGluLeuPheTrpThrSerAspPro 808
QY 4617 ACCTGCTATACCTCAGATGGGAAGGAATCCTCCCCACCAAAAGAGGGTTAGATATGTC 4676
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
809 LeuProTyrrPhePheGluGlyLysIleAepLeuThrProGluGluGlyLysPheVal 828
QY 4677 CAACAGATACATCGTCTAACCCACCTAGAGAACTAAACACCTGCAGCAGTGGTCAGAACAA 4736
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
829 LysGlyLeuHisGlnPheThrHisLeuGlyValGluLysMetMetArgLeuIleLysLys 848
QY 4737 TCCCTTATCATCTTCGAGGCTACCAGGATGGCTGACTCGGTGTGTCAAACATTTGTGTG 4796
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
849 SerArgTyrrGlnValProAenLeuLysSerValAlaGlnLysIleAenSerCysLys 868
QY 4797 CCCTGCCAGCTGTTAATGCTAATCCTCCAGATGCTCCAGGGAAGAGACTAAGGGGA 4856
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
869 AlaCysAlaPheThrAenAlaThrLysThrTyrrLysGluProGlyLysArgGlnAArgLys 888
QY 4857 AGCCACCCAGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGCTAAATACGGA 4916
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
889 AspArgProGlyValTyrrTrpGluValAspPheThrGluValLysProGlyMetTyrrGly 908
QY 4917 AACAAATACCTATGTTTTTTAGACACCTTTTTCAGGATGGGTAGAGCTTATCTTACT 4976
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
909 AsnLysTyrrLeuLeuValPheValAspThrPheSerGlyTyrrValGluAlaPheProThr 928
QY 4977 AAGAAAGAGACTTCAACCGTGGTGGCTTAAATAATACTGGAAGAAATTTTCCAGATTT 5036
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
929 LysThrGluThrAlaGlnIleValAlaLysLysIlePheGluGluIleLeuProArgTyrr 948
QY 5037 GGAATACCTAAGGTAAATAGGTGAGACAATGGTCCAGCTTTTGTGCCAGGTAAGTCAG 5096
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
949 GlyValProLysValIleGlySerAspAsnGlyProAlaPheValAlaGlnValSerGln 968
QY 5097 GGACTGGCCCAAGATATTGGGGATTGATTGGAATCGATTGTGCATACAGACCCCAAGC 5156
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
969 GlyLeuAlaThrGlnLeuGlyIleAspTrpLysLeuHisCysAlaTyrrArgProGlnSer 988
QY 5157 TCAGGACAGCTAGAGGATGATAGACCATTAAGAGACCTTACTTAATTTGACCGCG 5216
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
989 SerGlyGlnValGluArgMetAenArgThrLeuLysGluThrLeuThrLysLeuAlaMet 1008
QY 5217 GAGACTGGCGTTAATGATTGATAGCTCTCCTGCCCTTTGTGCTTTTGTAGGGTTAGGAAC 5276
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1009 GluThrGlyLysAspTrpValAlaLeuLeuProLeuAlaLeuPheArgAlaAArgAsn 1028
QY 5277 ACCCTGGACAGTTTGGCTGACCCCTATGAATTAATCTACGGGGAGACCCCCCATTTG 5336
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1029 ThrProGlyArgPheGlyLeuThrProPheGluValLeuTyrrGlyGlyProProLeu 1048
QY 5337 GTAGAAATCTCTCT-----GTACATAGTGTGAGTGTGCTTCTCCAGCCTTTGTTC 5390
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1049 IleLysAspGlyGlyThrLeuValProAspSerGlySerValLeuProSerSerLeuLeu 1068
QY 5391 TCTAGGCTCAAGCAGCTTGTAGTGGTGAGACAACAGCGGTGGAGCAACTCCGGGAGGCC 5450
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1069 IleHisLeuLysAlaLeuLysValIleArgThrGlnIleTrpAspGlnLeuLysThrAla 1088
QY 5451 TACTCAGGAGGAGAGACTTTCAGATCCCATCGTTTCCAACTGGGAGATTCAGTCTAC 5510
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1089 TyrrThr---ProGlyThrThrAlaValProHisGluPheGlnValGlyAspGlnValLeu 1107
QY 5511 GTTAGAGCCACCGTCAGCAAACTCGACACTCGGTGGAGGCGCTTATCTCGTACTT 5570
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1108 ValArgArgHisArgThrGlySerLeuGluProArgTrpLysGlyProTyrrLeuValLeu 1127
QY 5571 TTGACCACACCAACGGCTGTGAAAGTCGAAGGAATCTCCACCTGGATCTCCATGTCATCCCAC 5630
```

```
Db      1128 LeuThrThrProThrAlaValLysValAspGlyIleAlaSerTrpIleHisAlaSerHis 1147
QY 5631 GTTAAACCGCGCCACCTCCCGAT-----TCGGGGTGGAAAGCCGAAAG 5675
Db      1148 ValLysArgAlaProSerGlnAspGluThrHisGluAspAsnTrpAlaValGluAla 1167
QY 5676 ACTGAAATCCCTTAAGCTTCGCCCTCCATCGCGTGGTTCCT 5717
Db      1168 ThrAspAsnProLeuLysLeuArgLeuArgArgSerPro 1181

RESULT 10
AAR75189
ID AAR75189 standard; protein; 1196 AA.
XX
AC AAR75189;
XX
DT 27-AUG-2003 (revised)
DT 30-MAY-1996 (first entry)
XX
DE Osteoinductive retrovirus RFB-14 pol gene product.
XX
KW RFB retrovirus; gag; pol; env; osteogenesis; osteoinductive protein;
KW bone development; osteoporosis; gene therapy; polymerase;
KW reverse transcriptase.
XX
OS Murine leukemia virus.
XX
PN DE4411718-Al.
XX
PD 12-OCT-1995.
XX
PF 05-APR-1994; 94DE-04411718.
XX
PR 05-APR-1994; 94DE-04411718.
XX
PA (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.
XX
PI Schmidt J, Gimbel W, Strauss P, Erfle V, Pedersen PS, Pedersen L;
PI Oestergaard M;
XX
DR WPI; 1995-352078/46.
DR N-PSDB; AAQ94266.
XX
PT RFB-14 retrovirus genome - and prodn. of osteo-inductive proteins.
XX
PS Claim 14; Fig 1; 46pp; German.
XX
CC The full-length proviral genomic sequence of retrovirus RFB-14 has been
CC determined. The virus codes for an osteoinductive protein, although the
CC precise location of the coding region has not yet been identified. The
CC virus may be useful in gene therapy of bone growth disorders such as
CC osteoporosis. The present sequence is that of the viral pol gene product.
CC (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ Sequence 1196 AA;

Alignment Scores:
Pred. No.: 2,66e-294 Length: 1196
Score: 3981.00 Matches: 752
Percent Similarity: 76.8% Conservative: 168
Best Local Similarity: 62.8% Mismatches: 264
Query Match: 27.2% Indels: 14
DB: Gaps: 9

US-10-723-552-3 (1-8132) x AAR75189 (1-1196)
QY 2160 GGGAGACGGGTTCGACCCCTCCCGAGCCAGGGTAACCTTGAAGGTGGAGGGGCAA 2219
Db      1 GlyGlyGlnGlyGlnGluProProGluProArgIleThrLeuThrValGlyGln 20
QY 2220 CCAGTTGAGTTCCTGCTTCATACCGGAGCGAAACATTTCAGTGTACTACAGCCATTAGGA 2279
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Db	21	ProValThrPheLeuValAspThrClyAlaGlnHisSerValLeuThrGlnAsnProGly	40
QY	2280	AAACTAAAGATAAAAAATCTCGGTGTAGTGGTCCACAGGCGCAACAACTATCCATGG	2339
Db	41	ProLeuSerAspArgSerAlaTrpValGlnGlyAlaThrGlyGlyLysArgTyrArgTrp	60
QY	2340	ACTACCCGAAGAACAGTTGACTTGGAGTGGAGCGGTAAACCCACTCGTTTCTGGGTATA	2399
Db	61	ThrThrAspArgLysValHisLeuAlaThrGlyLysValThrHisSerPheLeuHisVal	80
QY	2400	CTGAGTCCCGCAGCACCCCTCTTAGGTAGACACTATTATGACCAAGATGGGGACACAATT	2459
Db	81	ProAspCysProTyrProLeuLeuGlyArgAspLeuLeuThrLysLeuValGlnLeu	100
QY	2460	TCCTTTTGAA---CAAGGGAAACACAGAGTGTCTCCAAATAACAACACTATCACTGTGTG	2516
Db	101	HisPheGluGlySerGlyAlaGlnValValGlyProLysGlyGlnProLeuGlnValLeu	120
QY	2517	ACCCTCCATTTAGATGACGAATATCGACTATACTCTCCCTAGTAAAGCCTGATCAAAAT	2576
Db	121	ThrLeuAsnLeuGluAspGluTyrArgLeuTyrGluThrSerAlaGluProGluAlaSer	140
QY	2577	ATA---CAATTCTCGTTGGACAGTTCCTCCCAAGCCTGGGCAGAAACCGCAGGGATGGT	2633
Db	141	ProGlySerThrTrpLeuSerAspPheProGlnAlaTrpAlaGluThrGlyGlyMetGly	160
QY	2634	TTGCAAGACGAAGTTCCCCCAAGATTATTCAACTGAAGGCCAGTGCGCACACAGTGTCA	2693
Db	161	LeuAlaValArgGlnAlaProLeuLeuLeileProLeuLysAlaThrSerThrProValSer	180
QY	2694	GTCCAGACAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTCGGCGCGCATGTCCAAAGA	2753
Db	181	IleLysGlnTyrProMetSerGlnGluAlaArgLeuGlyIleLysProHisIleGlnArg	200
QY	2754	TTAATCCACACAGGCGATCCTAGTTCCTGCTCCAACTCCCTGGATATCTCCCTGCTACCG	2813
Db	201	LeuLeuAspGlnGlyIleLeuValProCysGlnSerProTrpAsnThrProLeuLeuPro	220
QY	2814	GTTAGAAGCCTGGGACTAATGACTATCGACAGTACAGGACTTGAGAGAGTCAATAAA	2873
Db	221	ValLysLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsnLys	240
QY	2874	CGGTGCAGGATATACACCCAACAGTCCCGAACCTTTATACCTCTTGTGTGCTCTCCCA	2933
Db	241	ArgValGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerGlyLeuPro	260
QY	2934	CCCCAACGGAGCTGGTATACAGTATTGGACTTAAAGGATCCCTTCTTCTCGCTGAGATT	2993
Db	261	ProSerHisArgTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeu	280
QY	2994	CACCCCACTAGCCAAACCACTTTTTCCTTCGATGGAGAGATCCAGGTACGGCAAGAAC	3053
Db	281	HisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyMetGlyIleSer	300
QY	3054	GGGCGACTCACCTGGACCCGACTGCCCCAAAGGTTCAAGAACTCCCGACCATCTTTGAC	3113
Db	301	GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAsp	320
QY	3114	GAAGCCCTACACAGAGACCTGGCCCACTTCAGGATCCAAACCCCTCAGGTGACCTCTCT	3173
Db	321	GluAlaLeuHisArgAspLeuAlaGlyPheArgIleGlnHisProAspLeuLeuLeuLeu	340
QY	3174	CAGTACGTGGATGACCTGCTTCTTGGCGGGAGCCACAAACAGGACTGCTTTAGAAGGCAC	3233
Db	341	GlnTyrValAspAspLeuLeuLeuAlaAlaThrSerGluLeuAspCysGlnGlnGlyThr	360
QY	3234	AAGCACTACTGCTGGAAATTGTCTGACCTTAGGCTACAGAGCCCTCTGCTAAGAAGGCCAG	3293
Db	361	ArgAlaLeuLeuGlnThrLeuGlyAspLeuGlyTyrArgAlaSerAlaLysLysAlaGln	380
QY	3294	ATTTCAGGAGAGAGGTAACTACTTGGGGTACAGTTTGGCGGACGGCAGCGATCGCTG	3353
Db	381	IleCysGlnGlnValLysTyrLeuGlyTyrLeuLeuLysGluGlyGlnArgTrpLeu	400

QY	3354	ACGAGGACGGAAGAAACATGTAGTCCAGATACCGGCCCCCAACACACAGCCAAACAATG	3413
Db	401	ThrGluAlaArgLysGluThrValMetGlyGlnProIleProLysThrProArgGlnLeu	420
QY	3414	AGCAGTTTTTTGGGACAGCTGGATTTTGACAGACTGTGGATCCCGGGTTTGGCACCTTA	3473
Db	421	ArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGluMet	440
QY	3474	GCAGCCCACTTACCCGCTAAACCAAGAAAAAGGGAAATTCCTCTGGGCTCTGAGCAC	3533
Db	441	AlaAlaProLeuTyProLeuThrLysThrGlyThrLeuPheAsnTrpGlyProAspGln	460
QY	3534	CAGAAGGCAATTGATGCTATCAAAAGGCCCTGCTGAGCGCACTGCTCTGGGCCCTCCCT	3593
Db	461	GlnLysAlaTyGlnGluIleLysGlnAlaLeuLeuThrAlaProAlaLeuGlyLeuPro	480
QY	3594	GACGTAACTAAACCTTTACCTTTATGTGTGATGAGGTAAAGGAGTAGCCCGGGAGTT	3653
Db	481	AspLeuThrLysProPheGluLeuPheValAspGluLysGlnGlyTyAlaLysGlyVal	500
QY	3654	TTAACCACCAACCTTAGGACCATGGAGAACCTGTGCCTCACTGTCAAAAGAAGCTCGAT	3713
Db	501	LeuThrGlnLysLeuGlyProTrpArgArgSerValAlaTyxLeuSerLysLysLeuAsp	520
QY	3714	CCTGTAGCCAGTGGTGGCCCATATGCTGAAGGCTATCGCAGCTGTGGCCATCACTGGTC	3773
Db	521	ProValAlaAlaGlyTrpProCysLeuArgMetValAlaAlaIleAlaValLeuThr	540
QY	3774	AAGGACCTGACAAATTGACTTTGGGACAGATAATTACTGTATAGCCCCCATGCAATTG	3833
Db	541	LysAspAlaGlyLysLeuThrMetGlyGlnProLeuValIleLeuAlaProHisAlaGlu	560
QY	3834	GAGAACTCTTCGGGACGCCCCAGACCGATGATGACCAACGCCCGCATGACCCACTAT	3893
Db	561	GluAlaLeuValLysGlnProProAspArgTrpLeuSerAsnAlaArgMetThrHisLysr	580
QY	3894	CAAAGCCTGCTTCTC---ACAGAGAGGGTCACGTTTCGCTCCACGACGCGCTCTCAACCC	3950
Db	581	GlnAlaMetLeuLeuAspThrAspArgValGlnPheGlyProValValAlaLeuAsnPro	600
QY	3951	GCACACTTCTGCTGAGAGACTGTATGAACAGTGACTCATCATGATGGCCATCAACTATTG	4010
Db	601	AlaThrLeuLeuPro---LeuProGluGluGlyAlaProHisAspCysLeuGluIleLeu	619
QY	4011	ATTGAGAGACTGGGTCGCAAGGACCTTACAGACATACCGCTGACTGGGAAGTGCTA	4070
Db	620	AlaGluThrHisGlyThrArgProAspLeuThrAspGlnProIleProAspAlaAspHis	639
QY	4071	ACCTGGTTCACTGACGGAAGCAGCTATGTGGTGGAAAGGTAAGAGGATGGCTGGGGGGCG	4130
Db	640	ThrTrpTySerAspGlySerPheLeuGlnGluGlyGlnArgLysAlaGlyAlaAla	659
QY	4131	GTGGTGACGGGACCCGACAGATCTGGCCAGCAGCGCTGCGGAGAAAGAACTTCAGACAA	4190
Db	660	ValThrThrGluThrGluValIleTrpAlaArgAlaLeuProAlaGlyThrSerAlaGln	679
QY	4191	AAGGCTGAGCTCATGCCCTCAGCAGCTTTTCGGCTGCGCCGAGGGAAATCCATAAC	4250
Db	680	ArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLysArgLeuAsn	699
QY	4251	ATTATACGGACAGCAGGTATGCTTTTGGCAGCTGCACGCTACATGGGGCCATCTATAA	4310
Db	700	ValTyThrAspSerArgTyAlaPheAlaThrAlaHisIleHisGlyGluIleTyArg	719
QY	4311	CAAAGGGGTTGCTTACCTCAGCAGGGAGGGAATAAAGAAACAAAGAGGAAATTCCTAAG	4370
Db	720	ArgArgGlyLeuLeuThrSerGluGlyArgGluIleLysAsnLysSerGluIleLeuAla	739
QY	4371	CTATTAGAGCCCTACATTTACAAAAGGCTAGCTATTATTATACATCTGCTCGGACATCAG	4430
Db	740	LeuLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysLeuGlyHisGln	759

```
QY 4431 AAAGCTAAAGATCTCATATCCAGAGGAAACACAGATGGCTGACCGGGTTGCAACAGCAGCA 4490
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
760 LysGlyAspSerAlaGluAlaArgGlyAsnArgLeuAlaAspGlnAlaAlaArgGluAla 779
QY 4491 GCC-----CAGGGTGTAACTCTTCGCTATAATAGAAATGCCCAAGCCCCA 4538
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
780 AlaIleLysThrProProAspThrSerThrLeuLeuIleGluAspSerThrProTyrThr 799
QY 4539 GAACCCAGACAGATGACACCCCTAGAACAGCTGGCAGAGATATAAAAGATAGACAGTTC 4598
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
800 ProAlaTyrPheHisTyrThrGluThrAspLeuLysLysLeuArgGluLeuGlyAlaThr 819
QY 4599 TCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATCCTGCCCCCAAA 4658
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
820 TyrAsnGlnSerLysGly----TyrTrpValPheGlnGlyLysProValMetProAspGln 838
QY 4659 GAAGGGTTAGATATATGCCAACAGATACATCGTCTAACCCACCTAGGAACATAAACACCTG 4718
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
839 PheValPheGluLeuLeuAspSerLeuHisArgLeuThrHisLeuGlyTyrGlnLysMet 858
QY 4719 CAGCAGTTGTC-----AGAACATCCCTTATCATGTTCTGAGGCTACAGGAGTGGCT 4772
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
859 LysAlaLeuLeuAspArgGlyGluSerProTyrTyrMetLeuAsnArgAspLysThrLeu 878
QY 4773 GACTCGGTGTCAAACATTGTCGCTGCAGCTGTTAATGCTAATCTTCCAGAATG 4832
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
879 GlnTyrValAlaAspSerCysThrValCysAlaGlnValAsnAlaSerLysAlaLysIle 898
QY 4833 CCTCCAGGGAAGACTAAAGGGAAGCCACCCAGCGCTCAGTGGGAAGTGGACTTCACCT 4892
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
899 GlyAlaGlyValArgValArgGlyHisArgProGlySerHisTrpGluIleAspPheThr 918
QY 4893 GAGGTAAAGCGGCTAAATACGGNAACAATACCTATTGTTTGTAGACACCTTTTCA 4952
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
919 GluValLysProGlyLeuTyrGlyTyrLysTyrLeuLeuValPheValAspThrPheSer 938
QY 4953 GGATGGGTAGAGGCTTATCTACTAAGAAAGACTTCAACCGTGGTGGCTAAAGAAATA 5012
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
939 GlyTrpValGluAlaPheProThrLysArgLysGluThrAlaArgValValSerLysLysLeu 958
QY 5013 CTGGAAGAAATTTTCAAGATTGGAACTCAAGTAATAGGGTCAGACAATGGTCCA 5072
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
959 LeuGluGluIlePheProArgPheGlyMetProGlnValLeuGlySerAspAsnGlyPro 978
QY 5073 GCTTTGTCGCCAGTAACTCAGGACTCGGCCAGATATGGGATTGATTCGAACTG 5132
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
979 AlaPheThrSerGlnValSerGlnSerValAlaAspLeuLeuGlyIleAspTrpLysLeu 998
QY 5133 CATTGTGCATACAGACCCCAAGCTCAGGACAGCTAGAGAGGATGAATAGAACATTAAA 5192
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
999 HisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLys 1018
QY 5193 GAGACCCCTTAAATGACCGGAGACTGGCGTTAATGATTGGATAGCTCTCTCGCCC 5252
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1019 GluThrLeuThrLysLeuThrLeuAlaGlyThrArgAspTrpValLeuLeuLeuPro 1038
QY 5253 TTTGTGCTTTTATAGGTAGAACACCCCTGGACAGTTTGGCTACGCCCTATGATTA 5312
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1039 LeuAlaLeuTyrArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGluIle 1058
QY 5313 CTCTACGGGGACCCCCCATTTGGTAGAAATGCTTCTGTACATAGTGTGAGCTGCTG 5372
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1059 LeuTyrGlyAlaProProProLeuValAsnPheHisAspProAspMetSerGluLeuThr 1078
QY 5373 CTTTCCGAGCCTTTGTTCTTAGCTCAAGGCATCTGAGTGGGTGAGACAACAGCGGTGG 5432
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1079 AsnSerProSerLeuGlnAlaHisLeuGlnAlaLeuGlnThrValGlnArgGluIleTrp 1098
QY 5433 AGCAACTCCGGAGGCTTACTAGGAGGAGGACATGTCAG---ATCCACATCGTTTC 5489
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1099 LysProLeuAlaGluAlaTyrArgAspGlnLeuAspGlnProValIlePheProPhe 1118
QY 5490 CAAGTGGGAGATTTCAGTCTACGTTAGACGCCACCGTGCAGGAAACCTCGAGTTCGGTGG 5549
```

```
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1119 ArgIleGlyAspSerValTrpValArgHisGlnThrLysAsnLeuGluProArgTrp 1138
QY 5550 AGGGCCCTTATCTCTACTTTTGACCACACACGCTGTGAAAGTCGAGGAATCTCC 5609
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1139 LysGlyProTyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleSer 1158
QY 5610 ACCTGATCCATGCATCCACGCTTAACCGGG-----CCACCTCCCGATTCCGGGTGG 5663
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1159 AlaTrpIleHisAlaAlaHisValLysAlaAlaThrThrProProIleLysProSerTrp 1178
QY 5664 AAAGCCGAAAAGACTGAAAATCCCTTAAGCTTCGCTCCATCGCTGGTTCCT 5717
Db ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1179 ArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGlyAlaPro 1196

RESULT 11
ADH76470
ID ADH76470 standard; protein; 1199 AA.
XX
AC ADH76470;
XX
DT 15-APR-2004 (first entry)
XX
DE Murine leukemia virus pol gene derived protein.
XX
KW chimeric plasmid; replicative retroviral genome; gag; pol;
KW murine leukemia virus; MLV; env; gibbon ape leukemia virus; GaLV; virion;
KW MLV-GaLV-type; gene therapy; pAM.
XX
OS Murine leukemia virus.
XX
PN FR2832424-A1.
XX
PD 23-MAY-2003.
XX
PF 20-NOV-2001; 2001FR-00014976.
XX
PR 20-NOV-2001; 2001FR-00014976.
XX
PA (GENE-) GENETHON III.
PI Audit M, Cosset FL;
XX
DR WPI; 2003-471779/45.
XX
PT Chimeric plasmid containing replicative retroviral genome, useful for
PT making positive control virions in testing for replication-competent
PT retrovirus.
PS Claim 4; SEQ ID NO 2; 70pp; French.
XX
CC The invention relates to a novel chimeric plasmid comprising a
CC replicative retroviral genome. The replicative retroviral genome
CC comprises the gag and pol sequences from a murine leukemia virus (MLV);
CC and a chimeric env sequence comprising regions corresponding to parts of
CC the envelope derived from: an MLV genome; and a gibbon ape leukemia virus
CC (GaLV). Virions produced by expressing the viral genome of the chimeric
CC plasmid are useful as positive controls in a test for detection of
CC replication-competent retroviruses in preparations of MLV-GaLV-type
CC retroviral vectors. For example, to ensure that the MLV-GaLV-type
CC retroviral vectors, intended for gene therapy, have no capacity for
CC replication. This sequence represents the protein derived from an MLV pol
CC gene from a pAM plasmid of the invention.
XX
SQ Sequence 1199 AA;

Alignment Scores:
Pred. No.: 2 91e-294 Length: 1199
Score: 3980.50 Matches: 755
Percent Similarity: 76.2% Conservatives: 165
Best Local Similarity: 62.5% Mismatches: 257
Query Match: 27.2% Indels: 31
DB: Gaps: 11
```







[illegible]

```
Db 660 ValThrThrGluThrGluValIleTrrAlaArgAlaLeuProAlaGlyThrSerAlaGln 679
QY 4191 AAGCTGAGCTCATGGCCCTCAGCAAGCTTTGGGCTGCCGAAGGAAATCATAAAC 4250
Db 680 ArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLysArgLeuAsn 699
QY 4251 ATTATACGACAGCAGGTATGCTTTGCCACTGCACACATACATGGGCGCATCTAAA 4310
Db 700 ValTyrThrAspSerGlnTyrAlaPheAlaThrAlaHisIleHisGlyGluIleTyrArg 719
QY 4311 CAAAGGGGGTGTACTCTCAGCAGGAGGAGGAAATAAGCAACAAAGAGGAAATCTTAAGC 4370
Db 720 ArgArgGlyLeuLeuThrSerGluGlyArgGluIleLysAsnLysSerGluIleLeuAla 739
QY 4371 CTATTAGAACCGTACATTTACAAAAGAGTACTATTATACATCTGCTCGACATCAG 4430
Db 740 LeuLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysLeuGlyHisGln 759
QY 4431 AAGCTAAAGATCTCATATCCAGAGGAACAGATGGCTACCGGTTGCCAAGCAGGCA 4490
Db 760 LysGlyAspGlyAlaGluAlaArgGlyAsnArgLeuAlaAspGlnAlaAlaArgGluAla 779
QY 4491 GCC-----CAGGGTGTAACTCTCTGCTATATAGAAATGCCCAAGGCCCA 4538
Db 780 AlaIleLysThrProProAspThrSerThrLeuLeuIleGluAspSerThrProTyrThr 799
QY 4539 GAACCCAGACAGATACACCTTAGAAGATCGGCAAGAGATAAAAGATAGACAGTTC 4598
Db 800 ProAlaTyrPheHisTyrThrGluThrAspLeuLysLysLeuArgGluLeuGlyAlaThr 819
QY 4599 TCTGAGACTCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATCCTGCCCCACAA 4658
Db 820 TyrAsnGlnSerLysGly----TyrTrpValPheGlnGlyLysProValMetProAspGln 838
QY 4659 GAAGGGTGTAGATATGTCACACAGATACATCGTCTAACCCACCTTAGGAATAAACACCTG 4718
Db 839 PheValPheGluLeuLeuAspSerLeuHisArgLeuThrHisLeuGlyTyrGlnLysMet 858
QY 4719 CAGCAGTTGTGTC-----AGAACATCCCTTATCATGTTCTGAGGCTACAGGAGTGGCT 4772
Db 859 LysAlaLeuLeuAspArgGlyGluSerProTyrTyrMetLeuAsnArgAspLysThrLeu 878
QY 4773 GACTCGGTGTCAAACATTTGTGCTGCCCTGCAGCTGGTGTATCTAATCTTCCAGATG 4832
Db 879 GlnTyrValAlaAspSerCysThrValCysAlaGlnValAsnAlaSerLysAlaLysIle 898
QY 4833 CCTCCAGGGAAGAGACTAAAGGGAAGCCACCCAGGCGCTCACTGGGAAGTGGACTTCAC 4892
Db 899 GlyAlaGlyValArgValArgGlyHisArgProGlySerHisTrpGluIleAspPheThr 918
QY 4893 GAGGTAAAGCGCGCTAAATACGGAACAAATACCTATTGTTTTTGTAGACACTTTTCA 4952
Db 919 GluValLysProGlyLeuTyrGlyTyrLysTyrLeuLeuValPheValAspThrPheSer 938
QY 4953 GGTGGGTAGAGGCTTATCTTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAATAATA 5012
Db 939 GlyTrpValGluAlaPheProThrLysArgGluThrAlaArgValValSerLysLysLeu 958
QY 5013 CTGGAAGAAATTTTCCAGATTTGGAATAACCTAAGTAATAGGTCAGACAATGGTCCA 5072
Db 959 LeuGluGluIlePheProArgPheGlyMetProGlnValLeuGlySerAspAsnGlyPro 978
QY 5073 GCTTTTGTCCCGAGTAAGTACGGGACTCGGCCAAGATATTTGGGATTTGATTGGAAC 5132
Db 979 AlaPheThrSerGlnValSerGlnSerValAlaAspLeuLeuGlyIleAspTrpLysLeu 998
QY 5133 CATTTGTCATACAGACCCCAAGCTCAGGACGATAGAGAGGATGAATACAACTTAA 5192
Db 999 HisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLys 1018
QY 5193 GAGACCCCTTACTAAATTTGACC---GCGGAGACTGGCGTTAATGATTGATAGCTCTCCTG 5249
```

```
Db 1019 GluThrLeuThrLysLeuThrLeuAlaAlaProGlyThrArgAspTrpValLeuLeuLeu 1038
QY 5250 CCCTTTGTCTTTTATAGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCTATGAA 5309
Db 1039 ProLeuAlaLeuTyrArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGlu 1058
QY 5310 TTACTCTAGGGGACCCCCCATTTGGTAGAAATTTGCTTCTATACATAGTCTGACGTG 5369
Db 1059 IleLeuTyrArgAlaProProProLeuValAsnPheHisAspProAspMetSerGluLeu 1078
QY 5370 CTGCTTTCCAGCCTTTGTCTCTAGGCTCAAGCACTTGAAGTGGGTGAGAACACGAGCG 5429
Db 1079 ThrAsnSerProSerLeuGlnAlaHisLeuGlnAlaLeuGlnThrValGlnArgGluIle 1098
QY 5430 TGGAGCAACTCTCGGAGGCGCTTACTCAGGAGGAGGAGACTTGCGAG---ATCCCACATCGT 5486
Db 1099 TrpAsnProLeuAlaGluAlaTyrArgAspGlnLeuAspGlnProValIleProHisPro 1118
QY 5487 TTCCAAGTGGGAGATTCACTTACGTTAGACGCCACGCTGCAGGAACCTCGAGACTCGG 5546
Db 1119 PheArgIleGlyAspSerValTrpValArgArgHisGlnThrLysAsnLeuGluProArg 1138
QY 5547 TGAAGCGGCTTATCTCTGCTACTTTTGACCACACCAACCGCTGTGAAAGTCGAAGAAATC 5606
Db 1139 TrpLysGlyProTyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIle 1158
QY 5607 TCCACCTGGATCCATCCATCCAGCTTAAACCGGCG-----CCACCTCCCGATTCCGGG 5660
Db 1159 SerAlaTrpIleHisAlaAlaHisValLysAlaAlaThrThrProProIleLysProSer 1178
QY 5661 TGGAAAGCCGAAAGACTGAAATCCCTTAAAGCTTCGCTCCATCCGCTGGTTCCT 5717
Db 1179 TrpArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGlyAlaPro 1197

RESULT 13
AAB12994
ID AAB12994 standard; protein; 1199 AA.
XX
AC AAB12994;
XX
DT 30-NOV-2000 (first entry)
XX
DE MLV reverse transcriptase protein sequence SEQ ID #41.
XX
DE Reverse transcriptase; RT; polymerase chain reaction; PCR; retrovirus;
XX
DE Reverse transcriptase; RT; polymerase chain reaction; PCR; retrovirus;
XX
DE Reverse transcriptase; RT; polymerase chain reaction; PCR; retrovirus;
XX
DE Murine leukemia virus.
XX
DE W0200042199-A1.
XX
PD 20-JUL-2000.
XX
PF 14-JAN-2000; 2000WO-US0000896.
XX
PR 15-JAN-1999; 99US-0116099P.
XX
PA (MOLE-) MOLECULAR BIOLOGY RESOURCES.
XX
PI Swaminathan N;
XX
PF 2000-482830/42.
XX
PT Novel genes encoding reverse transcriptase polypeptides modified by
PT altering or adding the integrase domains by truncation internally and/or
PT at the C-termini, useful in cDNA synthesis and amplification procedures.
XX
PS Claim 1; Page 154-159; 189pp; English.
XX
CC This invention relates to isolated polynucleotide sequences encoding a
CC polypeptide with RNA dependent DNA polymerase (reverse transcriptase RT)
CC activity. RTs are found in a variety of retroviruses and their defining
CC activity is the ability to synthesise a cDNA strand using an RNA
```

CC template. The invention includes sequences AAA87808-A87822 and AAA87840-A87842 which represent nucleotide sequences encoding RT polypeptides. CC Included in the invention are reverse transcriptase protein sequences. CC AAB12989-B12995. Sequences AAA87829-A87839 and AAA87843-A87897 represent CC oligonucleotides used in the identification and synthesis of the RT CC nucleotide sequences of the invention. Many of the RT nucleotide CC sequences encode modified RT proteins, which exhibit improved stability CC and/or improved solubility, relative to naturally occurring reverse CC transcriptases. The modified RT nucleotide and protein sequence are used CC in improved polymerase chain reaction (PCR) methods. The nucleotide CC sequences can be used in sequencing methods. The present sequence CC represents a reverse transcriptase protein of the invention

XX Sequence 1199 AA;

SQ

Alignment Scores:

Pred. No.:	2,01e-293	Length:	1199
Score:	3969.50	Matches:	752
Percent Similarity:	76.3%	Conservative:	170
Best Local Similarity:	62.2%	Mismatches:	254
Query Match:	27.1%	Indels:	33
DB:	3	Gaps:	12

US-10-723-552-3 (1-8132) x AAB12994 (1-1199)

QY 2160 GGGAGACGGGTTTCGGACCCCTCCCGAGCCCGAGGTAACCTTTGAAGTGGAGGGCAA 2219  
 DB 1 GlyGlyGlnGlnGlnProProGluProArgIleThrLeuLysValGlyGlyGln 20

QY 2220 CCAGTTGAGTTCTGTTGATACCGGAGGAAACATTAGTGTCTACTAGCCATTAGGA 2279  
 DB 21 ProValThrPheLeuValAspThrGlyAlaGlnHisSerValLeuThrGlnAsnProGly 40

QY 2280 AAAATAAAGATAAATAATCTCGGGTGTGGTGGCCACAGGSCAACACAGTATCCATGG 2339  
 DB 41 ProLeuSerAspLysSerAlaIleValGlnGlyAlaThrGlyLysArgIleThrArgIle 60

QY 2340 ACTACCGAAGAACAGTTGACTTGGAGTGGGAGCGGTAACCCACTCGTTTCTGTGTCATA 2399  
 DB 61 ThrThrAspArgLysValHisLeuAlaThrGlyLysValThrHisSerPheLeuHisVal 80

QY 2400 CTTGAGTCCGAGCACCCCTCTTAGTAGAGACTTATTGACCAAGATGGGAGACACAATT 2459  
 DB 81 ProAspCysProTyrrProLeuLeuGlyArgAspLeuLeuThrLysLeuLysAlaGlnIle 100

QY 2460 TCTTTTGA--CAAGGGAAACAGAAAGTGTGCAATAAACAACCTATCACTGTGTG 2516  
 DB 101 HisPheGluGlySerGlyAlaGlnValMetGlyProMetGlyGlnProLeuGlnValLeu 120

QY 2517 ACCCTCCAAATGATGACGAATATCGACTATATCTCTCCCTAGTAAAGCCTGATCAAAAT 2576  
 DB 121 ThrLeuAsnIleGluAspGluHisArgLeuLeuHisGluThrSerLysGluProAspValSer 140

QY 2577 ATA---CAATTCTGTTGAAACAGTTTCCCAAGCTGGGAGAAACCCGAGGATGGGT 2633  
 DB 141 LeuGlySerThrTrpLeuSerAspPheProGlnAlaTrpAlaGluThrGlyGlyMetGly 160

QY 2634 TTGGCAAGCAAGTTCCCCCAACAAGTTATTCAACTGAAGCCAGTGCACACACAGTCA 2693  
 DB 161 LeuAlaValArgGlnAlaProLeuIleIleProLeuLysAlaThrSerThrProValSer 180

QY 2694 GTCAGACAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTCGGCCGCGATCTCCAAAGA 2753  
 DB 181 IleLysGlnTyrrProMetSerGlnGluAlaArgLeuGlyIleLysProHisIleGlnArg 200

QY 2754 TTAATCCACAGGGATCTAGTCTCTGTGTCATCTCCCTGGAATCTCCCTGCTACCG 2813  
 DB 201 LeuLeuAspGlnGlyLeuValProCysGlnSerProTyrrAsnThrProLeuLeuPro 220

QY 2814 GTTAGAAGCCTGGGACATGACTATCGACAGTACAGGACTTCAGAGAGTCAATAAA 2873  
 DB 221 ValLysLysProGlyThrAsnAspTyrrArgProValGlnAspLeuArgGluValAsnLys 240

QY 2874 CGGGTCAGGATATACACCCAAACAGTCCGAAACCTTTATACCTCTGTGTCTCTCCCA 2933  
 DB 241 ArgValGluAspIleHisProThrValProAsnProTyrrAsnLeuLeuSerGlyLeuPro 260

QY 2934 CCCCACGAGCTGGTATACAGTATTGGACTTAAGGATGCCTTCTCTCTCCCTGAGATTA 2993  
 DB 261 ProSerHisGlnTrpTyrrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeu 280

QY 2994 CACCCCACTAGCAACCACTTTTTCGCTTCGAATGGAGAGATCCAGGTACGGGAACAAC 3053  
 DB 281 HisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGluMetGlyIleSer 300

QY 3054 GGGCAGCTCCTCGGACCCGACTGCCCAAGGTTCAAGAACTCCCGACCATCTTTGAC 3113  
 DB 301 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAsp 320

QY 3114 GAAGCCCTACACAGAGACTGCGCACTTCAGATTCAGATTCACACCTCAGGTGACCTCTC 3173  
 DB 321 GluAlaLeuHisArgAspLeuAlaAspPheArgIleGlnHisProAspLeuIleLeuLeu 340

QY 3174 CAGTACGTGGATGACTGCTCTCTGCGGGAGCCACCAACACGAGTCTTAGAAGGCACG 3233  
 DB 341 GlnTyrrValAspLeuLeuAlaThrSerGlnLeuAspCysGlnGlnGlyThr 360

QY 3234 AAGGCACTACTGCTGGAATTTGCTACCTAGCTACAGAGCCTCTCTTAAGAAGGCCACG 3293  
 DB 361 ArgAlaLeuLeuGlnThrLeuGlyAsnLeuGlyTyrrArgAlaSerAlaLysLysAlaGln 380

QY 3294 ATTTCGAGGAGAGAGTAACTACTTGGGTACAGTTTTCGGGAGCGGAGGAGGTGGCTG 3353  
 DB 381 IleCysGlnLysGlnValLysTyrrLeuGlyTyrrLeuLeuLysGluGlyGlnArgTrpLeu 400

QY 3354 ACGGAGGCGACGGAAGAAATCTAGTCCAGATACCGGCCCAACACACGACCAACAAATG 3413  
 DB 401 ThrGluAlaArgLysGluThrValMetGlyGlnProProLysThrProArgGlnLeu 420

QY 3414 AGAGAGTTTTGGGACAGCTGGATTTTCAGACTGTGGATCCCGGGTTTCGACCTTA 3473  
 DB 421 ArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGluMet 440

QY 3474 CGAGCCCACTCTACCCGCTAACCAAGAAAGGGGAATTTCTCTGGGCTCTGAGACAC 3533  
 DB 441 AlaAlaProLeuTyrrProLeuThrLysThrGlyThrLeuPheAsnTrpGlyProAspGln 460

QY 3534 CAGAAGGCATTTGATGCTATCAAAAGCCCTGCTGAGCGCACCTGCTGCGCCCTCCCT 3593  
 DB 461 GlnLysAlaTyrrGlnGluIleLysGlnAlaLeuThrAlaProAlaLeuGlyLeuPro 480

QY 3594 GACGTAACCTAAACCCCTTTTACCTTTTATGTGGATGAGCGTAAAGGAGTAGCCCGGGAGTT 3653  
 DB 481 AspLeuThrLysProPheGluLeuPheValAspGluLysGlnGlyTyrrAlaLysGlyVal 500

QY 3654 TTAACCCAAACCTTAGGACCATGAGAGACCTGTGCGCTACTCTCAAAAGAGCTCGAT 3713  
 DB 501 LeuThrGlnLysLeuGlyProTrpArgProValAlaTyrrLeuSerLysLysLeuAsp 520

QY 3714 CCTGTAGCCAGTGTGGCCCATATGCTCGTGAAGGCTATCGCAGCTGTGCGCATCTGTC 3773  
 DB 521 ProValAlaAlaGlyTyrrProProCysLeuArgMetValAlaAlaIleAlaValLeuThr 540

QY 3774 AAGCAGCTGACAAATTGACTTTTGGGACAGAAATAAATCTGTAATAGCCCCCATGATTG 3833  
 DB 541 LysAspAlaGlyLysLeuThrMetGlyGlnProLeuValIleLeuAlaProHisAlaVal 560

QY 3834 GAGAACATCGTTGGCAGCCCCCAGACCGATGATGACCAACCGCCGATGACCCACTAT 3893  
 DB 561 GluAlaLeuValLysGlnProProAspArgTrpLeuSerAsnAlaArgMetThrHisTyrr 580

QY 3894 CAAGCCCTGCTCTC---ACAGAGAGGCTCAGTTCGCTCCGACCGCGCTCTCAACCT 3950  
 DB 581 GlnAlaLeuLeuLeuAspThrAspArgValGlnPheGlyProValValAlaLeuAsnPro 600

QY 3951 GCCACTCTCTCGCTGAAGAGACTGATGAACCAAGTGAATGATGATGATGATGATGATGAT 4010

Db	601	AlaThrLeuLeuPro---LeuProGluGluGlyLeuGlnHisAsnCysLeuAspIleLeu	619
Qy	4011	ATTGAGGAGACTGGGGTCCCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGCTA	4070
Db	620	AlaGluAlaHisGlyThrArgProAspLeuThrAspGlnProLeuProAspAlaAspHis	639
Qy	4071	ACCTGGTTCTACTGACGGAAGCAGCTATCTGCTGGAAGTAAAGAGATGGCTGGGGCGCG	4130
Db	640	ThrTrpTyThrAspGlySerSerLeuLeuGlnGluGlyGlnArgLysAlaGlyAlaAla	659
Qy	4131	GTGGTGACGGGACCCGACGATCTGGCCAGACGACCTGCGGGAAGAACTTCAGCACAA	4190
Db	660	ValThrThrGluThrGluValIleTrpAlaLysAlaLeuProAlaGlyThrSerAlaGln	679
Qy	4191	AAGCTCAGCTCATGGCCCTCAGCAAGCTTTCGGCTGCGGCAAGGGAATCCATAAAC	4250
Db	680	ArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLysLeuAsn	699
Qy	4251	ATTATACGACAGCAGCTATGCCCTTTGCGACTGCACACGTACATGGGGCCATCTATAA	4310
Db	700	ValTyThrAspSerArgTyAlaPheAlaThrAlaHisIleHisGlyGluIleTyArg	719
Qy	4311	CAAAAGGGGTGCTTACTCAGCAGGAGGGAATAAAGAACAAAGAGGAAATCTTAAGC	4370
Db	720	ArgArgGlyLeuLeuThrSerGluGlyLysGluIleLysAsnLysAspGluIleLeuAla	739
Qy	4371	CTATTAGACCCGTACATTTTACCAAAAGGCTAGCTATTATACACTGCTCTGCACATCAG	4430
Db	740	LeuLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysProGlyHisGln	759
Qy	4431	AAAGCTAAAGATCTCATATCCAGAGGAAACACAGATGGCTCACCGGGTTGCCAAGCAGGA	4490
Db	760	LysGlyHisSerAlaGluAlaArgGlyAsnArgMetAlaAspGlnAlaLaaArgLysAla	779
Qy	4491	GCCAGGGGTGTAACTTCTGCTTATTAATAAGAAATGCCCAAAGCC-----	4535
Db	780	Ala-----IleThrGluThrProAspThrSerThrLeuLeuIle	792
Qy	4536	-----CCAGAACCAGAGGACAGTACACCCCTAGAAGACTGGCAAGAG	4577
Db	793	GluAsnSerSerProTyThrThrSerGluHisPheHisTyThrValThrAspIleLysAsp	812
Qy	4578	ATAAAAAGATAGACACG---TTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGAT	4634
Db	813	LeuThrLysLeuGlyAlaIleTyAspLysThrLysLysTyTrpValTyr-----Gln	830
Qy	4635	GGGAAGAAATCTTGCCCCCAAAAGAGGTGTAGAATATGTCCAACAGATACATCGTCTA	4694
Db	831	GlyLysProValMetProAspGlnPheThrPheGluLeuLeuAspPheLeuHisGlnLeu	850
Qy	4695	ACCCACCTAGGAACATAACACCTGCACAGTGTGTGTCAGACA-----TCCCTTATCAT	4748
Db	851	ThrHisLeuSerPheSerLysMetLeuAlaLeuLeuGluArgSerHisSerProTyTrp	870
Qy	4749	GTTCCTGAGGCTACCAGGAGTGGCTGACTCGTGTGTCAACATTGTGTGCTCGCAGCTG	4808
Db	871	MetLeuAsnArgAspArgThrLeuLysAsnIleThrGluThrCysLysAlaCysAlaGln	890
Qy	4809	GTTAATGCTAATCTCCAGAAATGCCCTCCAGGGAAGAGACTAAAGGGGAAGCCACCAGGC	4868
Db	891	ValAsnAlaSerLysSerAlaValLysGlnGlyThrArgValArgGlyHisArgProGly	910
Qy	4869	GCTCAGTGGGAATGGACTTCATCAGAGTTAAAGCCGGCTAAATACGGAACAATACTACCTA	4928
Db	911	ThrHisTrpGluIleAspPheThrGluIleLysProGlyLeuTyGlyTyLysTyLeu	930
Qy	4929	TTGGTTTTCTAGACACTTTTACAGATGGGTAGAGGCTTATCTCTACAGAAAGAGACT	4988
Db	931	LeuValPheIleAspThrPheSerGlyTyTrpIleGluAlaPheProThrLysLysGluThr	950
Qy	4989	TCAACCCGTGGTGGCTAAAAAATACTCGAAGAAATTTTCCAGATTTGGAATACCTAAG	5048

Db	951	AlaLysValValThrLysLysLeuLeuGluGluLeuPheProArgPheGlyMetProGln	970
Qy	5049	GTAATAGGCTCAGACAAATGCTCCAGCTTTTGTGGCCAGCAGTAAGTCAGGACTGGCCCAAG	5108
Db	971	ValLeuGlyThrAspAsnGlyProAlaPheValSerLysValSerGlnThrValAlaAsp	990
Qy	5109	ATATTGGGAGTATTGGAAACTGCATTGTGCATACAGACCCCAAGACTCAGACAGGTA	5168
Db	991	LeuLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnVal	1010
Qy	5169	GAGAGATGAATAGAACCATTAAGAGACCCCTTACTAAATTGACCCGGAGACTGGCGTT	5228
Db	1011	GluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThrLeuAlaThrGlySer	1030
Qy	5229	AATGATTGGATAGCTCTCCTGCCCTTTGTGCTTTTGTAGGTTAGGAACACCCCTGCAGAC	5288
Db	1031	ArgAspTrpValLeuLeuLeuProLeuAlaLeuTyrArgAlaArgAsnThrProGlyPro	1050
Qy	5289	TTTGGGCTGACCCCTATGAATTACTCTACGGGGGACCCCCCATTTGGTGTAGAAATTGCT	5348
Db	1051	HisGlyLeuThrProTyrGluIleLeuTyrGlyAlaProProProLeuValAsnPhePro	1070
Qy	5349	TCGTACATAGTCTGACGTGCTGCTTTCCACGCTTTTGTCTCTAGGCTCAAGGACATT	5408
Db	1071	AspProAspMetThrArgValThrAsnSerProSerLeuGlnAlaHisLeuGlnAlaLeu	1090
Qy	5409	GAGTGGGTGAGACAACGAGCGTGGAGCAACTCCGGGAGGCCTACTCAGGAGGAGGAC	5468
Db	1091	TyrLeuValGlnHisGluValTrpArgProLeuAlaAlaTyrGlnGluGlnLeuAsp	1110
Qy	5469	TTGCAG---ATCCCCACATCGTTTCCAAAGTGGGAGATTTCAGTCTACGTTAGACGCCACCGT	5525
Db	1111	ArgProValValProHisProTyrArgValGlyAspThrValTrpValArgArgHisGln	1130
Qy	5526	GCAGGAAACCTCGAGACTCGGTGGAGGGCCCTTACTCTGTACTTTTGACCAACCAACG	5585
Db	1131	ThrLysAsnLeuGluProArgTrpLysGlyProTyrThrValLeuLeuThrThrProThr	1150
Qy	5586	GCGTGGAAGTCCGAAGGAATCTCCACCTGGATCCATGCATCCCACTTAAACCGCGCCA	5645
Db	1151	AlaLeuLysValAspGlyIleAlaAlaTrpIleHisAlaAlaHisValLysAlaAlaAsp	1170
Qy	5646	ACCT-----CCCATTTCG-----GGGTGGAAGCCGAAAGACTGAAATCCCTT	5690
Db	1171	ProGlyGlyGlyProSerSerArgLeuThrTrpArgValGlnArgSerGlnAsnProLeu	1190
Qy	5691	AAGCTTCGCTCCATCGCGTGGTTCCT	5717
Db	1191	LysIleArgLeuThrArgGluAlaPro	1199
RESULT 14			
AA17947			
ID	AA17947 standard; protein; 1224 AA.		
XX	AA17947;		
XX	03-AUG-1999 (first entry)		
DT	MoMLV pol gene product.		
DE	MoMLV pol gene product.		
XX	Viral packaging signal; chimeric; type C retrovirus; gag gene; MoMLV;		
KW	murine VL30; retroviral vector; biotechnology; pharmaceutical;		
KW	gene therapy; Moloney murine leukemia virus; pol gene.		
XX	Moloney murine leukemia virus.		
OS	Moloney murine leukemia virus.		
XX	WO925862-A2.		
PN	27-MAY-1999.		
PD	19-NOV-1998; 98WO-US024667.		
XX	19-NOV-1997; 97US-0066148P.		
PR			

XX (NATU-) NATURE TECHNOLOGY INC.  
 PA Hodgson CP, Zink MA, Guoping X;  
 PI WPI: 1999-347491/29.  
 XX N-PSDB; AAX77178.  
 XX New chimeric packaging signal useful in gene therapy.  
 PT Disclosure; Fig 4; 61pp; English.  
 XX  
 XX The invention relates to a new chimeric viral packaging signal that  
 CC comprises: (i) essential packaging nucleic acid sequence, from a  
 CC mammalian type C retrovirus, functionally joined to (ii). (ii) at least  
 CC one non-essential packaging nucleic acid sequence (lacking the gag gene)  
 CC from murine VL30. The chimeric viral packaging signals are used to  
 CC produce retroviral vectors for transfer of genetic material in gene or  
 CC cell therapy, or other biotechnological applications. Typical uses are in  
 CC production of pharmaceuticals (e.g. enzymes, antibiotics, antibodies,  
 CC antisense RNA, cytokines etc.) in animals or cell cultures. Elimination  
 CC of the gag gene results in safer vectors (reduced formation of  
 CC replication competent retrovirus as a result of recombination events  
 CC involving the gag gene). Also vectors that contain the chimeric packaging  
 CC signal have high infectious titers (about 1 million transduction units  
 CC (TU)/ml), comparable with that for gag-containing vectors and higher than  
 CC for known gag-free vectors. The present sequence represents the pol gene  
 CC product of the MoMLV virus  
 XX  
 SQ Sequence 1224 AA;

## Alignment Scores:

Pred. No.: 3,7e-292 Length: 1224  
 Score: 3953.00 Matches: 750  
 Percent Similarity: 74.6% Conservative: 170  
 Best Local Similarity: 60.8% Mismatches: 256  
 Query Match: 27.0% Indels: 58  
 DB: 2 Gaps: 11

US-10-723-552-3 (1-8132) x AAX17947 (1-1224)

QY 2160 GGGAGACGGGTTCGGAGCCCTCCCGAGCCCGAGTAACTTTGAAGTGGAGGGGCA 2219  
 DB 1 GlyGlyGlnGlyGlnGluProProGluProArgGileThrLeuLysValGlyGln 20  
 QY 2220 CCAGTTGAGTTCCTGGTTCATACCGGAGCAACATTCAGTCTACTACAGCCATTAGGA 2279  
 DB 21 ProValThrPheLeuValAspThrGlyAlaGlnHisSerValLeuThrGlnAsnProGly 40  
 QY 2280 AAACATAAGATATAAATTCCTGGGTGATGGGTGCCACAGGCAACACAGTATCCATGG 2339  
 DB 41 ProLeuSerAspLysSerAlaTrpValGlnGlyAlaThrGlyLysArgTyrArgTrp 60  
 QY 2340 ACTACCCGAGAACAGTTCAGTTCGGAGTGGGAGCGGTAAACCCACTCGTTTCGTGCATA 2399  
 DB 61 ThrThrAspArgLysValHisLeuAlaThrGlyLysValThrHisSerPheLeuHisVal 80  
 QY 2400 CTGAGTGCACGACCCCTCTAGGTAGAGACTTATTGACCAAGTGGGAGCACAAATT 2459  
 DB 81 ProAspCysProTyrProLeuLeuGlyArgAspLeuLeuThrLysLeuLysAlaGlnLe 100  
 QY 2460 TCTTTTGA---CAAGGGAAACACAGAGTGTCTGCAAAATAACACCTATCACTGTGTG 2516  
 DB 101 HisPheGluGlySerGlyAlaGlnValMetGlyProMetGlyGlnProLeuGlnValLeu 120  
 QY 2517 ACCCTCCAAATAGATGCAATATACGACTATACTCTCCCTAGTAAAGCTGATCAAAAT 2576  
 DB 121 ThrLeuAsnIleGluAspGluHisArgLeuHisGluThrSerLysGluProAspValSer 140  
 QY 2577 ATA---CAATTCTGCTGGAACAGTTTCCCAAGCTCGGACAGAACCCGAGGATGGGT 2633  
 DB 141 LeuGlySerThrTrpLeuSerAspPheProGlnAlaTrpAlaGluThrGlyGlyMetGly 160

QY 2634 TTGGCAAGCAAGTTTCCCCCAACAGTATTCACTGAAGCCAGTCCACACACAGTGTCA 2693  
 DB 161 LeuAlaValArgGlnAlaProLeuIleIleProLeuLysAlaThrSerThrProValSer 180  
 QY 2694 GTCAGACAGTACCCCTTGGATTAAGAGCTCAAGAGGAATTCGGCGCGCATGTCCAAAGA 2753  
 DB 181 IleLysGlnTyrProMetSerGlnGluAlaArgLeuGlyIleLysProHisIleGlnArg 200  
 QY 2754 TTAATCCAAACAGGCAATCCTAGTTCCTGCAATCTCCCTGGGAATATCTCCCTGCTACCG 2813  
 DB 201 LeuLeuAspGlnGlyIleLeuValProCysGlnSerProTrpAsnThrProLeuLeuPro 220  
 QY 2814 GTTAGAAACCTGGGACTAATGACTATCGACAGTACAGGACTTGAAGAGGTCAATAAA 2873  
 DB 221 ValLysLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsnLys 240  
 QY 2874 CGGGTGCAGGATATACACCAACAGTCCCGAACCTTATACTCTGTGTCTCTCCCA 2933  
 DB 241 ArgValGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerGlyLeuPro 260  
 QY 2934 CCCCAACGGAGCTGGTATACAGTATTGGACTTAAAGATGCTCTTCTGCTGAGATTA 2993  
 DB 261 ProSerHisGlnTrpTyrThrValLeuAspLeuLysAspAlaPheCysLeuArgLeu 280  
 QY 2994 CACCCACTAGCAACCACTTTTTCGCTTCGAATGGAGAGATCCAGGTACGGGAAGAAC 3053  
 DB 281 HisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGluMetGlyIleSer 300  
 QY 3054 GGGAGCTCCTGGACCCGACTGGCCCAAGGGTTCAAGAACTCCCCCGACCATCTTTGAC 3113  
 DB 301 GlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAsp 320  
 QY 3114 GAAGCCCTACACAGAGACTCGCCCAACTTTCAGATCCACACCCCTCAGTGCACCTCCTC 3173  
 DB 321 GluAlaLeuHisArgAspLeuAlaAspPheArgIleGlnHisProAspLeuLeuLeu 340  
 QY 3174 CAGTACGTGGATGACCTGCTCTCGCGGAGGACCAACACAGAGCTGTCTAGGAAGCACG 3233  
 DB 341 GlnTyrValAspAspLeuLeuAlaAlaThrSerGluLeuAspCysGlnGlnGlyThr 360  
 QY 3234 AAGGCACTACTCTGGAATTGCTGACCTAGCTACAGGCTCTGCTAAGAGGCCACG 3293  
 DB 361 ArgAlaLeuLeuGlnThrLeuGlyAsnLeuGlyTyrArgAlaSerAlaLysAlaGln 380  
 QY 3294 ATTTCCAGGAGAGGTAACTACTTGGGTACAGTTTTCGGGCGGAGGCGGAGTGGCTG 3353  
 DB 381 IleCysGlnLysGlnValLysTyrLeuGlyTyrLeuLeuLysGluGlyGlnArgTrpLeu 400  
 QY 3354 ACGGAGGCAACGGAAGAACTGTAGTCCAGATACCGGCCCCCAACACACCAACAAATG 3413  
 DB 401 ThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGlnLeu 420  
 QY 3414 AGAGAGTTTTGGGAGACAGTGGATTTTGCAGACTGTGGATCCCGGGGTTCGACCTTA 3473  
 DB 421 ArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGluMet 440  
 QY 3474 GCAGCCCACTCTACCCGCTAACCAAGAAAGGGAATTCCTCCTGGCTCCTGAGCAC 3533  
 DB 441 AlaAlaProLeuTyrProLeuThrLysThrGlyThrLeuPheAsnTrpGlyProAspGln 460  
 QY 3534 CAGAAGGCATTTGATGCTTATCAAAAGGCCCTGTGAGCGCACCTGCTGCGCCCTCCCT 3593  
 DB 461 GlnLysAlaTyrGlnGluIleLysGlnAlaLeuLeuThrAlaProAlaLeuGlyLeuPro 480  
 QY 3594 GACGTAACTAAACCTTTTACCTTTATGTGATGATGCGGTAAAGGAGTACCCCGGGAGTT 3653  
 DB 481 AspLeuThrLysProPheGluLeuPheValAspGluLysGlnGlyTyrAlaLysGlyVal 500  
 QY 3654 TTAACCCAAACCTTAGGACCATGGAGAGACTGTGCGCTACCTGCTCAAGAGTCCGAT 3713  
 DB 501 LeuThrGlnLysLeuGlyProTrpArgProValAlaTyrLeuSerLysLysLeuAsp 520  
 QY 3714 CTTGTAGCCAGTGGTGGCCCATATGCTGAAGGCTATCGCAGCTGTGGCCATCTGGTCT 3773

```
Db 521 ProValAlaAlaGlyTrpProProCysLeuArgMetValAlaAlaAlaValLeuThr 540
QY 3774 AAGGACGCTGACAAATGTACTTTGGGACAGAATATAACTGTATAGCCCCCATGCAATGG 3833
Db 541 LysAspAlaGlyLysLeuThrMetGlyGlnProLeuValIleLeuAlaProHisAlaVal 560
QY 3834 GAGAAATCGTTCCGAGCCCGACGAGCCGATGATGACCAACCCCGCATGACCCACTAT 3893
Db 561 GluAlaLeuValLysGlnProProAspA-gtrpLeuSerAsnAlaArgMetThrHisTyr 580
QY 3894 CAAAGCCTGCTTCTC---ACAGAGAGGGTCACGCTTCGCTCCACGACCGCTCTCAACCCCT 3950
Db 581 GlnAlaLeuLeuLeuAspThrAspArgValGlnPheGlyProValValAlaLeuAsnPro 600
QY 3951 GCACATCTTCTGCTCAGAGACTGATGAACAGTACTCATGATTTGCCATCACTATTGG 4010
Db 601 AlaThrLeuLeuPro---LeuProGluGluGlyLeuGlnHisAsnCysLeuAspIleLeu 619
QY 4011 ATTGAGGAGCTGGGTCCGACGAGGACCTTACAGACATACCGCTGACTGGAGAGTGCTA 4070
Db 620 AlaGluAlaHisGlyThrArgProAspLeuThrAspGlnProLeuProAspAlaAspHis 639
QY 4071 ACCTGCTTCACTACCGAAGCAGCTATGTGGTGAAGGTAAGAGGATGGCTGGGGCGCG 4130
Db 640 ThrTrpTyrThrAspGlySerLeuLeuGlnGluGlyGlnArgLysAlaGlyAlaAla 659
QY 4131 GTGGTGGACGGGACCCGACGATCTGGCGCAGCAGCTCGCGGAGGAACTTCAGACAA 4190
Db 660 ValThrThrGluThrGluValIleTrpAlaLysAlaLeuProAlaGlyThrSerAlaGln 679
QY 4191 AAGCTGAGCTCATGCGCTCAGCAGAGCTTTCGGCTGGCGGACCGGAAATCCATAAAC 4250
Db 680 ArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLysLysLeuAsn 699
QY 4251 ATTTATACGACAGCAGGTATGCTTTCGACTGCACACGTACATGGGGCATCTATAAA 4310
Db 700 ValTyrThrAspSerArgTyrAlaPheAlaThrAlaHisIleHisGlyGluIleTyrArg 719
QY 4311 CAAAGGGGTGTCTTACCTCAGCAGGGGAGGAAATAAGAACAAAGAGGAAATCTTAAGC 4370
Db 720 ArgArgGlyLeuLeuThrSerGluGlyLysGluIleLysAsnLysAspGluIleLeuAla 739
QY 4371 CTATTAGAGCCGTATATTACCAAAAGCTAGCTATTATACATGCTCTGACACATCAG 4430
Db 740 LeuLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysProGlyHisGln 759
QY 4431 AAAGCTAAAGATCTCATATCCAGGAAACACGATGGCTGACCGGTTGCCAAGCAGGCA 4490
Db 760 LysGlyHisSerAlaGluAlaArgGlyAsnArgMetAlaAspGlnAlaAlaArgLysAla 779
QY 4491 GCCAGGGGTGTTAACTTCTGCTTATAATAGAAATGCCAAAGCC----- 4535
Db 780 Ala-----IleThrGluThrProAspThrSerThrLeuLeuIle 792
QY 4536 -----CCAGAACCCAGCAGCAGTACACCTCAGAACGTCGCAAGAG 4577
Db 793 GluAsnSerSerProTyrThrSerGluHisPheHisTyrThrValThrAspIleLysAsp 812
QY 4578 ATAAAAAAGATAGACCAAG---TTCTCTGACATCCCGAGGGACCTGTCTATACCTCAGAT 4634
Db 813 LeuThrLysLeuGlyAlaIleTyrAspLysThrLysLysTyrTrpValTyr-----Gln 830
QY 4635 GGAAGGAATCTGCGCCCAAGAGGGTTAGAAATATGTCACAGATACATCTCTA 4694
Db 831 GlyLysProValMetProAspGlnPheThrPheGluLeuLeuAspPheLeuHisGlnLeu 850
QY 4695 ACCCATAGGAACCTAAACACCTGCGACGAGTTGGTCAGAAC-----TCCCTTATCAT 4748
Db 851 ThrHisLeuSerPheSerLysMetLysAlaLeuLeuGluArgSerHisSerProTyrTyr 870
QY 4749 GTTCTCAGGCTACAGGAGTGCTGACTCGGTGGTCAAAACATTTGTGGCCCTCCAGCTG 4808
Db 870
```

```
871 MetLeuAsnArgAspArgThrLeuLysAsnIleThrGluThrCysLysAlaCysAlaGln 890
QY 4809 GTAATGCTTAATCCTTCCAGAAATGCTCCAGGAAGAGACTAAGGGAAGCCACCCAGGC 4868
Db 891 ValAsnAlaSerLysSerAlaValLysGlnGlyThrArgValArgGlyHisArgProGly 910
QY 4869 GCTCACTGGGAAGTGGACTCTCACTGAGGTAAAGCCGGCTAAATACGGAACAAATACCTA 4928
Db 911 ThrHisTrpGluIleAspPheThrGluIleLysProGlyLeuTyrGlyTyrLysTyrLeu 930
QY 4929 TTGGTTTTGTAGACACCTTTTCAGATGGGTAGAGCTTATCTCTACTAAGAAAGAGACT 4988
Db 931 LeuValPheIleAspThrPheSerGlyTyrIleGluAlaPheProThrLysLysGluThr 950
QY 4989 TCAACCGTGGCTTAAAAAATACTCGGAAGAAATTTTCCAGATTTCCGAATACCTAAG 5048
Db 951 AlaLysValValThrLysLysLeuLeuGluIlePheProArgPheGlyMetProGln 970
QY 5049 GTAATAGGCTCAGACAATGCTCAGCTTTTGTGCCCAGCTAAGTCAGGACTGGCCAAAG 5108
Db 971 ValLeuGlyThrAspAsnGlyProAlaPheValSerLysValSerGlnThrValAlaAsp 990
QY 5109 ATATTGGGATTCATTGGAACTGTCATTCATACAGACCCCAAGCTCAGGACAGGTA 5168
Db 991 LeuLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnVal 1010
QY 5169 GAGAGATCAATAGAACCATTAAGAGACCTTACTAAATTCACCGCGGAGACTGGCGTT 5228
Db 1011 GluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThrLeuAlaThrGlySer 1030
QY 5229 AATGATGATGACTCTCTCGCTCCCTTTGCTTTTGGGTTAGGAACACCCCTCGACAG 5288
Db 1031 ArgAspTrpValLeuLeuLeuLeuProLeuAlaLeuTyrArgAlaArgAsnThrProGlyPro 1050
QY 5289 TTTGGCTGACCCCTTATGAATTTACTACGGGGGACCCCCCTTGTGTAAGAAATGTCT 5348
Db 1051 HisGlyLeuThrProTyrGluIleLeuTyrGlyAlaProProProLeuValAsnPhePro 1070
QY 5349 TCTGTACATAGTGTGACGTGCTGCTTTCAGCCCTTCTCTAGGCTCAAGGCACCT 5408
Db 1071 AspProAspMetThrArgValThrAsnSerProSerLeuGlnAlaHisLeuGlnAlaLeu 1090
QY 5409 GAGTGGTGGACAAACAGCAGCTGGAGGCAACTCCGGGAGGCTCTCTCAGGAGGAGGAC 5468
Db 1091 TyrLeuValGlnHisGluValTrpArgProLeuAlaAlaTyrGlnGlnLeuAsp 1110
QY 5469 TTGCAG---ATCCACATCGTTTCCAAAGTGGGAGATTCAGTCTACCTTAGACCCACCGT 5525
Db 1111 ArgProValValProHisProTyrArgValGlyAspThrValTrpValArgArgHisGln 1130
QY 5526 GCAGGAACTCGAGACTCGGTGGAGGGCCCTTATCTCTGTAATTTGACCAACCAAG 5585
Db 1131 ThrLysAsnLeuGluProArgTrpLysGlyProTyrThrValLeuLeuThrThrProThr 1150
QY 5586 GCTGTGAAAGTCGAAGGAATCTCCACTCGATCCATCCATCCACCTTAAACCGCGCCA 5645
Db 1151 AlaLeuLysValAspGlyIleAlaAlaTrpIleHisAlaAlaHisValLysAlaAlaAsp 1170
QY 5646 CTCCCGATTCG----- 5657
Db 1171 ProGlyGlyGlyProMetAlaArgSerThrLeuSerLysProLeuLysAsnLysValAsn 1190
QY 5658 -----GGGTGGAAAGCCGAAAG 5675
Db 1191 ProArgGlyProLeuLeuProLeuIleLeuSerSerArgLeuThrTrpArgValGlnArg 1210
QY 5676 ACTGAAATCCCTTAAAGCTTCGCTCCATCCATCGGTGGTTCCT 5717
Db 1211 SerGlnAsnProLeuLysIleArgLeuThrArgGluAlaPro 1224
Db 1224
```

RESULT 15  
ADS73450  
ID ADS73450 standard; protein; 744 AA.





```
Qy 1263 AACCATCCCTTCTCGAGGATCCCAACGCTCACGGGTTGGTGGAGTCCCTTATG 1322
Db |||||||
Qy 447 AsnHisProProPheSerGluAspProGlnArgLeuThrGlyLeuValGluSerLeuMet 466
Db |||||||
Qy 1323 TTCTCTCACAGCCTACTTGGGATGATTGTCAACAGCTGTGCAGACACTCTTCACAACC 1382
Db |||||||
Qy 467 PheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnThrLeuPheThrThr 486
Db |||||||
Qy 1383 GAGGAGCGAGAGAAATTCTATTAGAGCTAGAAAAAATGTTCTGGGGCCGACGGGCGA 1442
Db |||||||
Qy 487 GluGluArgGluArgIleLeuLeuGluAlaArgLysAsnValProGlyAlaAspGlyArg 506
Db |||||||
Qy 1443 CCACAGCGGTTGCNAATGAGATTGACATGGGATTTCCCTTAACCTGCCCGGTTGGGAC 1502
Db |||||||
Qy 507 ProThrArgLeuGlnAsnGluIleAspMetGlyPheProLeuThrArgProGlyTrpAsp 526
Db |||||||
Qy 1503 TACAACACCGCTGAAGGTAGGAGAGCTTCAAAATCTATCGCAGAGCTCTGGTGGCGGGT 1562
Db |||||||
Qy 527 TyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArgGlnAlaLeuValAlaGly 546
Db |||||||
Qy 1563 CTCGGGGCCCTCAAGACGGCCCACTAAATTGGCTAAGTAAGAGAAAGTGATGACGGGA 1622
Db |||||||
Qy 547 LeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysValArgGluValMetGlnGly 566
Db |||||||
Qy 1623 CCGAATGAACCCCTCTGTTTCTTGTAGAGGCTCTTGAAGCCTTCAGGCGGTACACC 1682
Db |||||||
Qy 567 ProAsnGluProProSerValPheLeuGluArgLeuLeuGluAlaPheArgArgTyrThr 586
Db |||||||
Qy 1683 CCTTTTTCATCCCTCAGAGGGCCCAAAAGCCTCAGTGGCTTTGGCTTTATAGACAG 1742
Db |||||||
Qy 587 ProPheAspProThrSerGluAlaGlnLysAlaSerValAlaLeuAlaPheIleGlyGln 606
Db |||||||
Qy 1743 TCAGCCTTGGATATTAGAAAGAGCTTCAGAGACTGGAAGGGTTACAGAGGCTGAGTTA 1802
Db |||||||
Qy 607 SerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGlyLeuGlnGluAlaGluLeu 626
Db |||||||
Qy 1803 CGTGATCTAGTGAAGGAGGAGAGAGATATATTACAAAGGGAGACAGAGAGAAAGG 1862
Db |||||||
Qy 627 ArgAspLeuValLysGluAlaGluLysValTyrTyrLysArgGluThrGluGluArg 646
Db |||||||
Qy 1863 GAACAAAGAAAAGAGAGAGAAAAGAGAGGAGGAGAAAGAGCTAATAACGGCAAGAG 1922
Db |||||||
Qy 647 GluGlnArgLysGluArgGluArgGluGluArgGluGluArgAsnLysArgGlnGlu 666
Db |||||||
Qy 1923 AAGAAATTGACTAAGATCTTGGCTGCTGAGTGGTTGAAGGGAAAAGCAATACGGAAGAG 1982
Db |||||||
Qy 667 LysAsnLeuThrLysIleLeuAlaAlaValValGluGlyLysSerAsnThrGluArgGlu 686
Db |||||||
Qy 1983 AGAGATTTTAGGAAAATTAGGTCAGGCCCTAGACAGTCAGGGAACCTGGGCAATAGGACC 2042
Db |||||||
Qy 687 ArgAspPheArgLysIleArgSerGlyProArgGlnSerGlyAsnLeuGlyAsnArgThr 706
Db |||||||
Qy 2043 CCACTCGACAGGACCAATGTGCATATTGTAAGAAAGAGACACACTGGGCAAGGAACCTGC 2102
Db |||||||
Qy 707 ProLeuAspLysAspGlnCysAlaTyrCysLysGluArgGlyHisTrpAlaArgAsnCys 726
Db |||||||
Qy 2103 CCCAAGAGGGAAACAAAGACCAAGGATCCTAGCTCTAGAGAAAGATAAAGAT 2156
Db |||||||
Qy 727 ProlLysLysGlyAenLysGlyProArgIleLeuAlaLeuGluAspLysAsp 744
Db |||||||
```

Search completed: February 14, 2006, 15:50:14  
Job time : 1473.8 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 15:55:36 ; Search time 42.0295 Seconds  
(without alignments)  
3199.272 Million cell updates/sec

Title: US-10-723-552-3  
Perfect score: 14636  
Sequence: 1 GCGTGTGTACGACTGTGG.....CTGTTTGCATCAAAAAA 8132

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DBV=xlp  
-Q=/abes/ABSSWEB spool/US10723552/runat\_14022006\_125146\_13079/app\_query.fasta\_1  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs802p -USER=US10723552 @CN 1 1 161 @runat\_14022006\_125146\_13079  
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5 COMB pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6 COMB pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H COMB pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PTUS COMB pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE COMB pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6099.5	41.7	1194	2	US-09-171-553B-5
2	5338	36.5	1737	2	US-09-309-572-13
3	5338	36.5	1737	2	US-09-171-553B-6
4	4329.5	29.6	1203	2	US-09-171-553B-13
5	3600	24.6	1079	1	US-09-075-272-4
6	3425	23.4	638	2	US-08-929-967-8
7	2866.5	19.6	660	2	US-09-111-085-2
8	2866.5	19.6	660	2	US-09-376-781-5
9	2640	18.0	524	2	US-09-171-553B-4
10	2500.5	17.1	665	1	US-08-929-967-7
11	2422	16.5	657	2	US-09-111-085-4
12	2422	16.5	657	2	US-09-376-781-4

13	2395	16.4	657	2	US-09-171-553B-10
14	2353	16.1	656	2	US-09-376-781-3
15	2353	16.1	656	2	US-09-171-553B-6
16	2297	15.7	1312	2	US-09-554-572-26
17	1796.5	12.3	622	2	US-09-075-272-2
18	1793	11.8	522	2	US-09-075-272-3
19	1505.5	10.3	667	2	US-09-315-127-5
20	1505.5	10.3	667	2	US-09-315-127-6
21	1505.5	10.3	667	2	US-09-070-630-13
22	1501.5	10.3	379	2	US-09-603-185-6
23	1499.5	10.2	768	2	US-08-979-847B-89
24	1477.5	10.1	673	2	US-09-075-272-5
25	1390.5	9.5	683	2	US-08-979-847B-198
26	1390.5	9.5	683	2	US-08-979-847B-200
27	1387.5	9.5	683	2	US-08-979-847B-208
28	1387.5	9.5	683	2	US-08-979-847B-210
29	1380.5	9.4	538	2	US-09-309-572-12
30	1380.5	9.4	538	2	US-09-718-096-12
31	1379.5	9.4	538	2	US-09-370-368-9
32	1331.5	9.1	654	2	US-08-979-847B-91
33	1320	9.0	665	2	US-09-309-572-14
34	1320	9.0	665	2	US-09-718-096-14
35	1318.5	9.0	632	2	US-09-315-127-2
36	1318.5	9.0	632	2	US-09-315-127-3
37	1303.5	8.9	768	2	US-09-120-653D-5
38	1277	8.7	645	2	US-09-315-127-8
39	1277	8.7	645	2	US-09-315-127-9
40	1262.5	8.6	654	2	US-09-315-127-11
41	1262.5	8.6	654	2	US-09-315-127-12
42	692.5	4.7	469	1	US-08-484-126-1
43	692.5	4.7	469	2	US-09-374-909-1
44	660.5	4.5	445	1	US-08-447-925-6
45	639.5	4.4	453	1	US-08-484-126-3

ALLEGMENTS

RESULT 1  
US-09-171-553B-5  
Sequence 5, Application US/09171553B  
Patent No. 6756227  
GENERAL INFORMATION:  
APPLICANT: GALBRAITH, DANIEL N.  
APPLICANT: HAWORTH, CHRISTINE  
APPLICANT: LEBES, GILLIAN M.  
APPLICANT: SMITH, KENNETH T.  
TITLE OF INVENTION: PORCINE RETROVIRUS  
FILE REFERENCE: CFV-5.01  
CURRENT APPLICATION NUMBER: US/09/171,553B  
CURRENT FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: PCT/GB97/01087  
PRIOR FILING DATE: 1997-04-18  
PRIOR APPLICATION NUMBER: GB 9702668.6  
PRIOR FILING DATE: 1997-02-10  
PRIOR APPLICATION NUMBER: GB 9608164.1  
PRIOR FILING DATE: 1996-04-19  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 1194  
TYPE: PRT  
ORGANISM: Porcine retrovirus  
US-09-171-553B-5

Alignment Scores:  
Pred. No.: 0  
Score: 6099.50  
Percent Similarity: 98.5%  
Best Local Similarity: 96.9%  
Query Match: 41.7%  
DB: 2  
Length: 1194  
Matches: 1154  
Conservative: 19  
Mismatch: 17  
Indels: 1  
Gaps: 1

US-10-723-552-3 (1-8132) x US-09-171-553B-5 (1-1194)

QY 2160 GGGAGCGGGTTCGGACCCCTCCCGAGCCAGCGTAACCTTTGAAGTGGAGGGCAA 2219  
DB |||||  
1 GlyArgArgGlySerAspProLeuProGluProArgValThrLeuLysValGluGln 20  
QY 2220 CCAGTTGAGTTCCTGGTTGATACCGAGCGCAACATTCAGTGTCTACTACAGCCATTAGGA 2279  
DB |||||  
21 ProValGluPheLeuValAspThrGlyAlaGluHisSerValLeuLeuGlnProLeuGly 40  
QY 2280 AAATAAGATAAAAATCCTGGGTGATGGGTGCCACAGGGCAACAAAGTATCCATGG 2339  
DB |||||  
41 LysLeuLysGluLysSerTrpValMetGlyAlaThrGlyGlnArgGlnTrpProTrp 60  
QY 2340 ACTACCCGAGAACAGTGTGCTGGAGTGGGAGCGGTAAACCACCTCGTTTCTGTGCATA 2399  
DB |||||  
61 ThrThrArgArgThrValAspLeuGlyValGlyArgValThrHisSerPheLeuValIle 80  
QY 2400 CTTGAGTGGCCAGCACCCCTCTTAGGTAGAGACTTATTGACCAGATGGGAGCACAAATT 2459  
DB |||||  
81 ProGluCysProValProLeuLeuGlyArgAspLeuLeuThrLysMetGlyAlaGlnIle 100  
QY 2460 TCTTTTGAACAGGGAAACAGAGTGTCTGCAATAACAAACCTATPACTGTGTGAAC 2519  
DB |||||  
101 SerPheGluGlnGlyArgProGluValSerValAsnAsnLysProIleThrValLeuThr 120  
QY 2520 CTCCAAATTAGATGACGAAATATGACTATCTCTCCCTAGTAAGCCTGATCAAAATATA 2579  
DB |||||  
121 LeuGlnLeuAspAspGluTrpArgLeuTrpSerProGlnValLysProAspGlnAspIle 140  
QY 2580 CAATTCTGGTTGGAACAGTTTCCCAAGCTGGGCGAGAAACCGCAGGAGTGGTTGGCA 2639  
DB |||||  
141 GlnSerTrpLeuGluGlnPheProGlnAlaTrpAlaGluThrAlaGlyMetGlyLeuAla 160  
QY 2640 AAGCAAGTTCCTCCACAAAGTTATTCAACTGAAGGCCAGTGCACACAGGTGTCAAGTCAGA 2699  
DB |||||  
161 LysGlnValProProGlnValIleGlnLeuLysAlaSerAlaThrProValSerValArg 180  
QY 2700 CAGTACCCCTTGAGTAAGAAGCTCAAGAAGAAATTCGGCGCGATGTCCTCAAGATTAAATC 2759  
DB |||||  
181 GlnTrpProLeuSerArgGluAlaArgGluGlyLeuTrpProHisValGlnArgLeuIle 200  
QY 2760 CAACAGGCACTCTAGTTCCTGTCCTCAATCTCCCTGGAAATCTCCCTGCTACCGGTAGA 2819  
DB |||||  
201 GlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLeuProValArg 220  
QY 2820 AAGCTGGGACTAATGACTATGACACAGTACAGACTTGAGAGAGTCAATAAACGGGTG 2879  
DB |||||  
221 LysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsnLysArgVal 240  
QY 2880 CAGGATATACCCCAAGTCCCGAACCTTATACCTCTGTGTGCTCTCCACCCCAA 2939  
DB |||||  
241 GlnAspIleHisProThrValProAsnProTyrAsnLeuLeuSerAlaLeuProProGlu 260  
QY 2940 CGGAGCTGGGTATACAGTATTGGACTTAAAGATGCCCTTCTTCTGCTCGAGATTACACCC 2999  
DB |||||  
261 ArgAsnTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisPro 280  
QY 3000 ACTAGCCAAACCACTTTTGGCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGSCAG 3059  
DB |||||  
281 ThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyThrGlyArgThrGlyGln 300  
QY 3060 CTACCTGGAGCCGACTGCGCCCAAGGTTCAAGAACTCCCGGACCATCTTTTGAGGAAGCC 3119  
DB |||||  
301 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePheAspGluAla 320  
QY 3120 CTACACAGACCTGGCCAACTTCAGGATCCAAACACCTCAGGTGACCTCTCCCTCCAGTAC 3179  
DB |||||  
321 LeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnTyr 340  
QY 3180 GTGGATGACCTGCTTCTGGCGGGAGCCCAACCAAGGACTGCTTTAGAAGGCACGAAGGCA 3239  
DB |||||  
341 ValAspAspLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGluGlyThrLysAla 360

QY 3240 CTACTGCTGGAATTGCTGACCTAGGCTACAGAGCCTCTGCTAAGAGGCCAGATTTC 3299  
DB |||||  
361 LeuLeuLeuGluLeuSerAspLeuGlyTyrArgAlaSerAlaLysLysAlaGlnIleCys 380  
QY 3300 AGGAGAGAGGTAAACATACCTTTGGGGTACAGTTTCGGGACGGGACGATGGCTGACGGAG 3359  
DB |||||  
381 ArgArgGluValThrTyrLeuGlyTyrSerLeuArgGlyGlyGlnArgTrpLeuThrGlu 400  
QY 3360 GCACGGAAGAAACTGTAGTCCAGATACCGGCCCCACACACCCCAACAATGAGAGAG 3419  
DB |||||  
401 AlaArgLysLysThrValValGlnIleProAlaProThrThrAlaLysGlnValArgGlu 420  
QY 3420 TTTTTCGGGACAGCTGATTTTCGACACTGTGTGATCCCGGGTTTCGACACTTAGCAGCC 3479  
DB |||||  
421 PheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaThrLeuAlaAla 440  
QY 3480 CCACTCTACCCCGCTAAACCAAGAAAAAGGGGAATTCCTCTGGGCTCCTGAGCACCAAG 3539  
DB |||||  
441 ProLeuTyrProLeuThrLysGluLysGlyGlyPheSerTrpAlaProGluHisGlnLys 460  
QY 3540 GCATTTGATGCTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCTGAGGTA 3599  
DB |||||  
461 AlaPheAspAlaIleLysLysAlaLeuLeuSerAlaProAlaLeuAlaLeuProAspVal 480  
QY 3600 ACTAAACCTTTTACCTTTATGTGGATGAGCGTAAGGAGTAGCCCGGGAGTTTAAAC 3659  
DB |||||  
481 ThrLysProPheThrLeuTyrValAspGluuArgLysGlyValAlaArgGlyValLeuThr 500  
QY 3660 CAAACCTTAGGACCATGGAAGACCTGTGCTCTACTCTGTCAAAAGAGCTCGATCTCTGTA 3719  
DB |||||  
501 GlnThrLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLeuAspProVal 520  
QY 3720 GCAGTGGTTGGCCATATGCTGAAGGCTATGCGAGCTGTGCGCCATACCTGCTCAAGGAC 3779  
DB |||||  
521 AlaSerGlyTrpProValCysLeuLysAlaIleAlaAlaValAlaIleLeuValLysAsp 540  
QY 3780 GCTGACAAATTGACTTTGGACAGATAATACTGTAATAGCCCCCATGATGAGGAAC 3839  
DB |||||  
541 AlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleAlaProHisAlaLeuGluAsn 560  
QY 3840 ATCGTTCCGCGAGCCCGACCGATGATGACCAACCGCCGCGATGACCCCATATCAAGC 3899  
DB |||||  
561 IleValArgGlnProProAspArgTrpMetThrAsnAlaArgMetThrHisTrpGlnSer 580  
QY 3900 CTGCTTCTACAGAGAGGTTCAGTTTCGTCACGAGCGCTCTCAACCTCGCATCTTT 3959  
DB |||||  
581 LeuLeuLeuThrGluArgValThrPheAlaProProAlaAlaLeuAsnProAlaThrLeu 600  
QY 3960 CTGCTGAAAGAGACTGATGAACCACTGACTCATGATTGCCATCAACTATTGATTGAGGAG 4019  
DB |||||  
601 LeuProGluGluThrAspGluProValThrHisAspCysHisGlnLeuLeuIleGluGlu 620  
QY 4020 ACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGTAACTGCTTTC 4079  
DB |||||  
621 ThrGlyValArgLysAspLeuThrAspIleProLeuThrGlyGluValLeuThrTrpPhe 640  
QY 4080 ACTGACGGAAGCACTATGTGTGGAGAGGTAAAGAGATGCTGGGGGGGGGGTGGAC 4139  
DB |||||  
641 ThrAspGlySerSerTyrValValGluGlyLysArgMetAlaGlyAlaAlaValValAsp 660  
QY 4140 GGGACCCGACGATCTGGGCGCAGCAGCTGCGGGAAGGAACCTTCAGCACAAAAGGCTGAG 4199  
DB |||||  
661 GlyThrArgThrIleTrpAlaSerSerLeuProGluGlyThrSerAlaGlnLysAlaGlu 680  
QY 4200 CTCATGGCCCTCACGCAAGCTTTTCGGGCTGGCCGGAAGGGAATCCATAAACATTATACG 4259  
DB |||||  
681 LeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluGlyLysSerIleAsnIleTyrThr 700  
QY 4260 GACAGAGGTATGCTTTGGGACTGGCACAGCTGATCGGGGCCATCTATAAACAAAGGGGG 4319  
DB |||||  
701 AspSerArgTyrAlaPheAlaThrAlaHisValHisGlyAlaIleTyrLysGlnArgGly 720  
QY 4320 TTGCTTACCTCAGCAGGGGGGAAATAAAGAAACAAAGAGGAAATTCCTAAGCCTATTAGAA 4379

Db 721 LeuLeuThrSerAlaGlyArgGluIleLysAsnLysGluGluIleLeuSerLeuLeuGlu 740  
QY 4380 GCGTATCATTTACCAAAAGCGTAGCTATTATATACACTGTCTCTGACATCAGAAAGCTAAA 4439  
Db 741 AlaLeuHisLeuProLysArgLeuAlaIleHisCysProGlyHisGlnLysAlaLys 760  
QY 4440 GATCTCATATCCAGAGAAACCGATGGCTTGACCGGTTCCCAAGCAGGCGAGCCCGGTT 4499  
Db 761 AspLeuIleSerArgGlyAsnGlnMetAlaAspArgValAlaLysGlnAlaGlnAla 780  
QY 4500 GTTAACTCTCTGCTATATAGAAATGCCCAAGAGCCCAAGACCCAGACGACAGTACACC 4559  
Db 781 ValAsnLeuLeuProIleIleGluThrProLysAlaProGluProArgArgGlnIleThr 800  
QY 4560 CTAGAAGACTGGCAAGAGATAAAAGATAGACAGTCTCTGAGACTCCCGGAAGGAGCC 4619  
Db 801 LeuGluAspTrpGlnGluIleLysLysIleAspGlnPheSerGluThrProGluGlyThr 820  
QY 4620 TGCTATACCTCAGATGGGAAGAAATCTCTGCCCAAGAAAGGGTTAGAAATATGCCAA 4679  
Db 821 CysTyThrSerTyrgLysGluIleLeuProHisLysGluGlyLeuGluTyValGln 840  
QY 4680 CAGATACATCGTCTAACCCACTAGCACTAAACACTGCAGCAGTGTGTCAGAACATCC 4739  
Db 841 GlnIleHisArgLeuThrHisLeuGlyThrLysHisLeuGlnGlnLeuValArgThrSer 860  
QY 4740 CTTTATCATGTTCTGAGGCTTACCAGGAGTGGCTGACTCGTGTGTCAAACTTGTGCC 4799  
Db 861 ProTyHisValLeuArgLeuProGlyValAlaAspSerValLysHisCysValPro 880  
QY 4800 TGCCAGCTGGTAACTCTTAACTCTCCAGAAATCCCTCAGGGAAGAGACTTAAGGGGAAGC 4859  
Db 881 CysGlnLeuValAsnAlaAsnProSerArgIleProProGlyLysArgLeuArgGlySer 900  
QY 4860 CACCCAGGCGCTCACTGGGAGTGGACTTCACTGAGGTAAAGCCGCTAAATACGAAAC 4919  
Db 901 HisProGlyAlaHisTrpGluValAspPheThrGluValLysProAlaLysTyrgLys 920  
QY 4920 AAATACCTATTGTTTGTAGACACTTTTTCAGGATGGTAGAGCTTCTACTAAG 4979  
Db 921 LysTyLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaTyProThrLys 940  
QY 4980 AAGAGACTTCAACCTGTGGCTAAATAAATACTCGAAGAAATTTTCCAGATTGGA 5039  
Db 941 LysGluThrSerThrValAlaLysLysIleLeuGluGluIlePheProArgPheGly 960  
QY 5040 ATACCTAAGGTAATAGGTCAGACATGTCTCAGCTTTTGTCCCGCAGTAAAGTACGGA 5099  
Db 961 IleProLysValIleGlySerAspAsnGlyProAlaPheValAlaGlnValSerGlnGly 980  
QY 5100 CTGGCCCAAGATATTGGGGATTGATTGAAACTGCTATTGTGCATACAGACCCCAAGCTCA 5159  
Db 981 LeuAlaLysIleLeuGlyIleAspTrpLysLeuHisCysAlaTyArgProGlnSerSer 1000  
QY 5160 GCACAGGTAGAGAGATGAATAGAACATTAAAGAGACCTTTACTAAATTTGACCGGGAG 5219  
Db 1001 GlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThrGlu 1020  
QY 5220 ACTGGGTTAATGATTGGATAGCTCTCTCGCTTGTGCTTTTGTAGGGTTAGGAACACC 5279  
Db 1021 ThrGlyIleAsnAspTrpMetAlaLeuLeuProPheValLeuPheArgValArgAsnThr 1040  
QY 5280 CTTGGACAGTTTGGGCTGACCCCTATGAATTAATCTACGGGGGACCCCGCCATTTGTA 5339  
Db 1041 ProGlyGlnPheGlyLeuThrProTyrgLysLeuLeuLeuTyrgLysProProLeuAla 1060  
QY 5340 GAAATGCTCTGTATCATAGTGTGACGTGCTGCTTCCAGCTTTGTCTCTAGGCTC 5399  
Db 1061 GluIleAlaPheAlaHisSerAlaAspValLeuLeuSerGlnProLeuPheSerArgLeu 1080  
QY 5400 AAGGCACTTGTGGGTGAGCAACGAGCGTGGAGGCAACTCCGGAGGCGCTTACTCAGGA 5459

Db 1081 LysAlaLeuGluTrpValArgGlnArgAlaTrpLysGlnLeuArgGluAlaTySer--- 1099  
QY 5460 GGAGGAGACTTGCAGATCCACATCGTTTCCAGTGGGAGATTCACTACGTTAGACGC 5519  
Db 1100 GlyGlyAspLeuGlnValProHisArgPheGlnValGlyAspSerValTyValArg 1119  
QY 5520 CACCGTGCAGGAAACCTCGAGACTCGGTGGAAGGCCCTTATCTCGTACTTTTGCACACA 5579  
Db 1120 HisArgAlaGlyAsnLeuGluThrArgTrpLysGlyProTyrgLysLeuValLeuThrThr 1139  
QY 5580 CCAACCGCTGTGAAGTCGAAAGAAATCTCCACTGCATCCATCCAGTCCCAACCG 5639  
Db 1140 ProThrAlaValLysValGluGlyIleProThrTrpIleHisAlaSerHisValLysPro 1159  
QY 5640 GCGCCACCTCCCGATTCGGGGTGAAGCCGAAAGAGCTGAAATCCCTTAAAGCTTCGC 5699  
Db 1160 AlaProProAspSerGlyTrpArgAlaGluLysThrGluAsnProLeuLysLeuArg 1179  
QY 5700 CTCCATCGCTGGTCTCTTACTCTCTGTCAATAAC 5732  
Db 1180 LeuHisArgLeuValProTyrgLysAsnAsn 1190  
RESULT 2  
US-09-309-572-13  
; Sequence 13, Application US/09309572  
; Patent No. 6440730  
; GENERAL INFORMATION:  
; APPLICANT: Heinrich-Pette-Institut  
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV  
; FILE REFERENCE: P50489  
; CURRENT APPLICATION NUMBER: US/09/309,572  
; CURRENT FILING DATE: 1999-05-11  
; EARLIER APPLICATION NUMBER: DE 198 56 463  
; EARLIER FILING DATE: 1998-11-26  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 1737  
; TYPE: PRT  
; ORGANISM: Moloney murine leukemia virus  
; FEATURE:  
; OTHER INFORMATION: gag-pol protein  
US-09-309-572-13  
Alignment Scores:  
Pred. No.: 0 Length: 1737  
Score: 5338.00 Matches: 1035  
Percent Similarity: 71.6% Conservative: 239  
Best Local Similarity: 58.2% Mismatches: 395  
Query Match: 36.5% Indels: 110  
DB: 2 Gaps: 26  
US-10-723-552-3 (1-8132) x US-09-309-572-13 (1-1737)  
QY 585 ATGGGACAGACGGTGACGACCCCTCTTAGTTTGTACTCTCGACCATTTGAGTAA 644  
Db 1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20  
QY 645 TCCAGGCTCATAAATTTGCTAGTTTACAGGTAGAGGGACCTTGGCAGACTTTCTGTGC 704  
Db 21 ArgIleAlaHisAsnGlnSerValAspValLysArgTrpValThrPheCysSer 40  
QY 705 TCTGAATGCGCCACATTCGATGTCGATGCGCATCAGAGGGACCTTTTAACTCTGAGATT 764  
Db 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60  
QY 765 ATCTCGGCTGTTAAAGCAGATTATTTTTCAGACTGACCGCGCTCTCATCCGATCAGGAG 824  
Db 61 IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80  
QY 825 CCTATATCTTACCTGCGCAAGATTGGCAGAGGATCTCCGCCATGGGTTAAACATGG 884  
Db 81 ProTyIleValThrTrpGluAlaLeuAlaPheAspProProTrpValLysProPhe 100

QY 885 CTGAATAAGCCAAAGAACCCAGGTCCCGAATTTCTGGCTCTTGGAGAGAAACAAACAC 944  
Db 101 ValHis---ProLysProProProProLeuProProSerAlaProSerLeuProLeuGlu 119  
QY 945 TCGGCTGAAGAACGTCAGCCCTCTCTCATATCTACCCCGAGATGAGGACCA---CCG 1001  
Db 120 ProProArgSerThrProProArgSerSerLeuTyrProAlaLeuThrProSerLeuGly 139  
QY 1002 GCTTGGCCGGAACCCCA-----TCT 1022  
Db 140 AlaLysProLysProGlnValLeuSerAspSerGlyGlyProLeuIleAspLeuLeuThr 159  
QY 1023 GTTCCCCACCCCTTATCTCGCACAGGNGCCCGAGGGACCCCTTGGCCCTCTCT--- 1079  
Db 160 GluAspProProProTyr-----ArgAspProArgProProProSer 173  
QY 1080 -----GGAGCTCCGCGGTGAGGGACCT----- 1103  
Db 174 AspArgAspGlyAsnGlyGlyGluAlaThrProAlaGlyGluAlaProAspProSerPro 193  
QY 1104 ---GTCAGGGACTCGAGCCGGAGGGGCCACCCCGGAGCGGACAGAGATCGCG 1160  
Db 194 MetAlaSerArgLeuArgGlyArgGluProProValAlaAspSerThrThrSerGln 213  
QY 1161 ACATTACCGCTGCGCAGTACGGCCCTCCACACCGGGGGCCAAATTGCAGCCCTCCAG 1220  
Db 214 AlaPheProLeuArgAlaGlyGly-----AsnGlyGln-----LeuGln 226  
QY 1221 TATTGGCCCTTTCTCTCGAGATCTCTATAATTGGAAACTAACCATCCCTCTTCTCG 1280  
Db 227 TyrTrpProPheSerSerSerAspLeuTyrAsnTrpLysAsnAsnAsnProSerPheSer 246  
QY 1281 GAGATCCCAACGCCTCACGGGGTGTGTGGAGTCCCTTANGTCTCTACACGCTTACT 1340  
Db 247 GluAspProGlyLysLeuThrAlaLeuIleGluSerValLeuIleThrHisGlnProThr 266  
QY 1341 TGGGATCATGTCAACAGCTGCTGCAGACACTTTCACACCCGAGGCGGAGAGAAATT 1400  
Db 267 TrpAspAspCysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGluLysGlnArgVal 286  
QY 1401 CTATTAGAGCTAGAAAAAATGTTCTTGGCGCCGACGGGCGACCCACGCGTTCGAAAAT 1460  
Db 287 LeuLeuGluAlaArgLysAlaValArgGlyAspAspGlyArgProThrGlnLeuProAsn 306  
QY 1461 GAGATTGACATGGGATTTCCCTTAACCTCGCCCGGTTGGGACTACAAACAGCGCTGAAGT 1520  
Db 307 GluValAspAlaAlaPheProLeuGluArgProAspTrpAspTyrThrThrGlnAlaGly 326  
QY 1521 AGGGAGAGCTGAAATCTATCCGAGGCTCTGTGCGGGTCTCGGGCGGCTCAAGA 1580  
Db 327 ArgAsnHisLeuValHisTyrArgGlnLeuLeuAlaGlyLeuGlnAsnAlaGlyArg 346  
QY 1581 CGGCCCACTAATTGGCTAAGTAAGAGAAAGTGTGAGGAGCCGAATGAACCCCTCT 1640  
Db 347 SerProThrAsnLeuAlaLysValLysGlyIleThrGlnGlyProAsnGluSerProSer 366  
QY 1641 GTTTTCTTGAGAGCTCTTGGAGCCCTTCAGCGGTACACCCCTTTTGTATCCACCTCA 1700  
Db 367 AlaPheLeuGluArgLeuLysGluAlaTyrArgArgTyrThrProTyrAspProGluAsp 386  
QY 1701 GAGGCCCAAAAGCCTCAGTGGCTTGGCCCTTTATAGGACAGTCAGCTTGGATATAGA 1760  
Db 387 ProGlyGlnGluThrAsnValSerMetSerPheIleTrpGlnSerAlaProAspIleGly 406  
QY 1761 AAGAAGCTTCAGAGCTGGAAGGTTACAGAGGCTCAGTTAGCTGTATCTAGTGAAGGAG 1820  
Db 407 ArgLysLeuGluArgLeuGluAspLeuLysAsnLysThrLeuGlyAspLeuValArgGlu 426  
QY 1821 GCAGAGAAAGTATATTCAAAAAGGAGACAGAGAGAAAGGAAACAAAGAAAGAGAGA 1880  
Db 427 AlaGluLysIlePheAsnLysArgGluThrProGluGluArgGluGluArgGluArg 446

QY 1881 GAAAGAGAGAAAGGAGGAAAGA-----CGT 1907  
Db 447 GluThrGluGluGlyGluArgArgThrGluAspGluGlnLysGluLysGluArg 466  
QY 1908 AATAAACGCAAGAGAAATTTGACTAAGATCTTGGCTGCAGCTGTTGAAGGAAAGC 1967  
Db 467 AspArgArgHisArgGluMetSerLysLeuLeuAlaThrValValSerGlyGlnLys 486  
QY 1968 AATACGGAAGAGAGAGATTTTAGAAAAATTAGTTCAGGCCTAGACAGTCAGGGAAC 2027  
Db 487 Gln-----AspArgGlnGlyGlu 493  
QY 2028 CTGGCAATAGGACCCCACTCGACAAAGACCAATGTGCATATTGTAAGAAAGACAGACAC 2087  
Db 494 ---ArgArgArgSerGlnLeuAspArgAspGlnCysAlaTyrCysLysGluLysGlyHis 512  
QY 2088 TGGCAAGGAACCTGCCCCAAGAG-----GGAACAAGAGCAAGG----- 2129  
Db 513 TrpAlaLysAspCysProLysLysProArgGlyProArgGlyProArgProGlnThrSer 532  
QY 2130 ATCTAGCTCTAGAACAAAGATAAAGATTAGGGAGACGGGTTCCGACCCCTCCCGAG 2189  
Db 533 LeuLeuThrLeuAspAsp-----GlyGlyGlnGlyGlnGluProProProGlu 548  
QY 2190 CCAGGCTAACTTTGAAGGTGGAGGGCAACCAAGTTGAGTTCCTGTTGATACCGGAGCG 2249  
Db 549 ProArgIleThrLeuLysValGlyGlyGlnProValThrPheLeuValAspThrGlyAla 568  
QY 2250 AAACATTAGTCTACTACAGCCATTAGGAAACTAAAGATAAAAATCTCGGTGATG 2309  
Db 569 GlnHisSerValLeuThrGlnAsnProGlyProLeuSerAspLysSerAlaTrpValGln 588  
QY 2310 GGTGCCACAGGCAACCAACGATATCCATCGACTACCCGAGACACTTGCTTTGGAGTG 2369  
Db 589 GlyAlaThrGlyGlyLysArgTyrArgTrpThrThrAspArgLysValHisLeuAlaThr 608  
QY 2370 GGAAGGCTAACCCACTCTTCTGTGTATACCTGAGTGCACGACCCCTCTTAGGTAGA 2429  
Db 609 GlyLysValThrHisSerPheLeuHisValProAspCysProTyrProLeuLeuGlyArg 628  
QY 2430 GACTTATTGACCAAGATGGGAGCACAAATTTCTTTTGA-----CAAGGGAAACCAAGATG 2486  
Db 629 AspLeuLeuThrLysLeuLysAlaGlnIleHisPheGluGlySerGlyAlaGlnValMet 648  
QY 2487 TCTGCAATAACAAACCTATCACTGTGTGACCTCCAATTAGATCAGCAATATCGACTA 2546  
Db 649 GlyProMetGlyGlnProLeuGlnValLeuThrLeuAsnIleGluAspGluHisArgLeu 668  
QY 2547 TACTCTCCCTAGTAAAGCTGATCAAAATATA---CAATTCTGTTGGAACAGTTTCCC 2603  
Db 669 HisGluThrSerLysGluProAspValSerLeuGlySerThrTrpLeuSerAspPhePro 688  
QY 2604 CAAGCTCTGGCAGAAACCCGAGGATGGGTTTGGCAAGCAAGTCTCCCCACAAGTTATT 2663  
Db 689 GlnAlaTrpAlaGluThrGlyMetGlyLeuAlaValArgGlnAlaProLeuIleIle 708  
QY 2664 CAACTGAGCCAGTCCCAACAGCTCAGTCAGTCAGACAGTACCCCTTGAGTAAAGAGCT 2723  
Db 709 ProLeuLysAlaThrSerThrProValSerIleLysGlnTyrProMetSerGlnGluAla 728  
QY 2724 CAAGAAGGAATTCGGCGCATGTCCAAAGATTAATCCACAGGGCATCTAGTTCTCTGTC 2783  
Db 729 ArgLeuGlyLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys 748  
QY 2784 CAATCTCCCTGGAAATCTCCCTGTCTACCGGTTAGAAAGCTTGGGACTAATGCTATCGA 2843  
Db 749 GlnSerProTrpAsnThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg 768  
QY 2844 CAGTACAGGACTTGAGAGAGGTCAATAACCGGTCAGGATATACACCAACAGTCCCG 2903  
Db 769 ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro 788  
QY 2904 AACCCCTTATAACCTCTTGTGTGCTCTCCCAACCCCAACGGAGCTGGTATACAGTATTGGAC 2963

Db 789 AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp 808  
QY 2964 TTAAGGATGCTTCTTCTGCTGAGATTACACCCACTAGCCACACACTTTTGGCTTC 3023  
Db 809 LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPhe 828  
QY 3024 GAATGGAGATCCAGGTACGGGAGAACCGGGGAGCTCACTGGAGCCCGACCTGCCCCAA 3083  
Db 829 GluTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln 848  
QY 3084 GGGTTCAAGAACTCCCGACCATCTTTGACGAGAGCCCTACACAGAGACCTGGCCAACTTC 3143  
Db 849 GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe 868  
QY 3144 AGGATCCAAACACCTCAGGTGACCCCTCTCCAGTACGTGATGACCTGCTTCTGGGGGA 3203  
Db 869 ArgIleGlnHisProAspLeuLeuLeuGlnTyrValAspAspLeuLeuAlaAla 888  
QY 3204 GCCACCAACAGACTGCTTAGAAGGACGACGACCTACTGCTGGAATGTCTGACCTA 3263  
Db 889 ThrSerGluLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAsnLeu 908  
QY 3264 GGTACAGAGCTCTGCTAAGAGGCCAGATTTGAGAGAGAGGTTCAGGAGAGGTAACTACTTGGG 3323  
Db 909 GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly 928  
QY 3324 TACAGTTTGGGACCGGACGCGATGCTGACGAGGACGAGGAGAACTGAGTCCAG 3383  
Db 929 TyrLeuLeuLysGlnGlyGlnArgTrpLeuThrGluAlaArgLysGlnThrValMetGly 948  
QY 3384 ATACCGGCCCAACACAGCAACAAATAGAGAGTTTTTGGGACAGCTGGATTTGC 3443  
Db 949 GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys 968  
QY 3444 AGACTGTGATCCCGGGTTGGACCTTAGGACGCGCCACTCTACCGCTAACCAAGAA 3503  
Db 969 ArgLeuTrpIleProGlyPheAlaGluMetAlaProLeuTyrProLeuThrLysThr 988  
QY 3504 AAAGGGGAATTCCTCGGGCTCTGAGCACCAAGAGCGCTTGTGATCTCAAAAAGGCC 3563  
Db 989 GlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTyrGlnGluLeuLysGlnAla 1008  
QY 3564 CTGCTGAGCGACCTGCTCTGGCCCTCCCTGACGTAACCTTACCTTTATGTG 3623  
Db 1009 LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal 1028  
QY 3624 GATGACGTAAGGAGTAGCCGGGAGTTTAAACCAACCTTAGGACCATGGAGAGA 3683  
Db 1029 AspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg 1048  
QY 3684 CTGTGCGCTACTCTGCAAGAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCCCTG 3743  
Db 1049 ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTrpProCysLeu 1068  
QY 3744 AAGGCTATCGAGCTGCGCCATCTGCTCAAGGACGCTGACAAATTTGATTTGGACAG 3803  
Db 1069 ArgMetValAlaAlaIleAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGln 1088  
QY 3804 ATATTAACGTATAGCCCCCATGATTTGGAGAACATCGTTGGGAGCCCCCAGACCGA 3863  
Db 1089 ProLeuValIleLeuAlaProHisAlaValGluAlaValLysGlnProProAspArg 1108  
QY 3864 TGGATGACCAACGCGCATGACCCACTATCAAGCTGCTTCTC---ACAGAGGGGTC 3920  
Db 1109 TrpLeuSerAsnAlaArgMetThrHisTyrGlnAlaLeuLeuAspThrAspArgVal 1128  
QY 3921 ACGTTGCTCCACGACCGCTCTCAACCTGCGACTCTTCTGCTGAAGAGACTGATGAA 3980  
Db 1129 GlnPheGlyProValValAlaLeuAsnProAlaThrLeuLeuPro---LeuProGluGlu 1147  
QY 3981 CCAGTGACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCGCGAAGGACCTT 4040  
Db 4041

1148 GlyLeuGlnHisAsnCysLeuAspIleLeuAlaGluAlaHisGlyThrArgProAspLeu 1167  
QY 4041 ACAGACATACCGCTGACTGGAGAAGTCTAACTGTTCTACTGACGGAACGACGATGTG 4100  
Db 1168 ThrAspGlnProLeuProAspAlaAspHisThrTrpTyrThrAspGlySerLeuLeu 1187  
QY 4101 GTGGAAGGTAAAGGATGCTGGGCGCGGTGTGGACGGGACCCGACGATCTGGGCC 4160  
Db 1188 GlnGluGlyGlnArgLysAlaGlyAlaAlaValThrThrGluThrGluValIleTrpAla 1207  
QY 4161 AGCAGCCTCGCGGAGGAACTTCAGCACAAAGCGCTGAGCTCATGGCCCTCAGCAAGCT 4220  
Db 1208 LysAlaLeuProAlaGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAla 1227  
QY 4221 TTGCGCTGCGCGGAGGAAATCCATAAATTTATACGACAGCAGGATGCTCTTTCGCG 4280  
Db 1228 LeuLysMetAlaGluGlyLysLysLeuAsnValTyrThrAspSerArgTyrAlaPheAla 1247  
QY 4281 ACTGCACAGTACATGGGCGCATCTATAAACAAGGGGTGCTTACCTCAGCAGGAGG 4340  
Db 1248 ThrAlaHisIleHisGlyGluIleTyrArgArgGlyLeuLeuThrSerGluGlyLys 1267  
QY 4341 GAAATAAGAACAAAGGAAATTTCTAAGCTATTAGAAGCCGTACATTTACCAAAAGG 4400  
Db 1268 GluIleLysAsnLysAspGluIleLeuAlaLeuLysAlaLeuPheLeuProLysArg 1287  
QY 4401 CTAGCTATTATACACTGCTCTGACATCAGAAAGCTAAAGATCTCATATCCAGAGAAAC 4460  
Db 1288 LeuSerIleIleHisCysProGlyHisGlnLysGlyHisSerAlaGluAlaArgGlyAsn 1307  
QY 4461 CAGATGGCTGACCGGGTTGCCAAGCAGGCGCCAGCGGTGTTAAACCTTCTGCTCATATA 4520  
Db 1308 ArgMetAlaAspGlnAlaAlaArgLysAlaAla-----IleThr 1320  
QY 4521 GAAATGCCCAAGGCC-----CCAGAACCCAGA 4547  
Db 1321 GluThrProAspThrSerThrLeuLeuIleGluAsnSerSerProTyrThrSerGluHis 1340  
QY 4548 CGCAGTACACCTAGACACTGGCAAGAGATAAAAAGATAGACCAG---TTCTCTGAG 4604  
Db 1341 PheHisTyrThrValThrAspIleLysAspLeuThrLysLeuGlyAlaIleTyrAspLys 1360  
QY 4605 ACTCGGAAGGACCTGTATACCTCAGATGGGAAGAAATCCTGCCCCCAAGAAAGGG 4664  
Db 1361 ThrLysLysTyrTrpValTyr-----GlnGlyLysProValMetProAspGlnPheThr 1378  
QY 4665 TTAGAATATGTCACACAGATACATGCTCTAACCCACTAGGAACCTAAACCTGCGACGAG 4724  
Db 1379 PheGluLeuLeuAspPheLeuHisGlnLeuThrHisLeuSerPheSerLysMetLysAla 1398  
QY 4725 TTGGTCAGAACAA-----TCCCTTATCATGTTCTGAGGCTACCAGAGTGGCTGACTCG 4778  
Db 1399 LeuLeuGluArgSerHisSerProTyrTyrMetLeuAsnArgAspArgThrLeuLysAsn 1418  
QY 4779 GTGGTCAAAACATGTGTGCCCTGCCAGCTGGTAAATGCTTAATCTTCCAGATGCCTCCA 4838  
Db 1419 IleThrGluThrCysLysAlaCysAlaGlnValAlaAlaSerLysSerAlaValLysGln 1438  
QY 4839 GGAAGAGACTAAGGGGAGCCAGCCAGCGCTCCTGGAAGTGTGACTTCACTAGGTA 4898  
Db 1439 GlyThrAlaValArgGlyHisArgProGlyThrHisTrpGluIleAspPheThrGluIle 1458  
QY 4899 AAGCGGCTAAATACGGAACAAATACCTATTGTTTGTAGACACCTTTTCAGGATG 4958  
Db 1459 LysProGlyLeuTyrGlyTyrLysTyrLeuLeuValPheIleAspThrPheSerGlyTrp 1478  
QY 4959 GTAGAGGCTTATCTACTAAGAAAGAGACTTCAACCGTGGTGGCTTAAATAATACTGGAA 5018  
Db 1479 IleGluAlaPheProThrLysLysGluThrAlaLysValValThrLysLysLeuLeuGlu 1498  
QY 5019 GAAATTTTCCAAAGATTGGGAATACCTAAGGTAATAGGTCAGACAATCGTCAGCTTTT 5078  
Db 1499 GluIlePheProArgPheGlyMetProGlnValLeuGlyThrAspAsnGlyProAlaPhe 1518

```
QY 5079 GTTCCAGGTAAGTCAGGACTGGCCCAAGATATTGGGATTTGATTTGGAACTGCATTGT 5138
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1538
1519 valserlyseValSerGlnThrValAlaAsePleuLeuGlyIleAspTrpLysLeuHisCys 1538
QY 5139 GCATACAGACCCCAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATTAAAGAGACC 5198
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1558
1539 AlaTyArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThr 1558
QY 5199 CTTACTAAATGACCGCGAGACTGGCGTTAATGATTGGATAGCTCTCTGCGCCTTTGTG 5258
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1578
1559 LeuThrLysLeuThrLeuAlaThrGlySerArgAspTrpValLeuLeuLeuProLeuAla 1578
QY 5259 CTTTTAGGGTTAGGAACCCCTGGACAGCTTTGGGCTGACCCCTATGAATACTCTAC 5318
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1598
1579 LeuTyArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyGluIleLeuTy 1598
QY 5319 GGGGGACCCCCCATTTGGTAGAAATTTGCTTCTGTACATAGTCTGACGTGCTTTCC 5378
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1618
1599 GlyAlaProProLeuValAsnPheProAspProAspMetThrArgValThrAsnSer 1618
QY 5379 CAGCCTTTGTTCTTAGGCTCAAGGCACCTTGAGTGGGTGAGACACAGAGCTGAGGCAA 5438
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1638
1619 ProSerLeuGlnAlaHisLeuGlnAlaLeuTyLeuValGlnHisGluValTrpArgPro 1638
QY 5439 CTCGGGAGGCTACTCAGGAGGAGGAGCTTCAG---ATCCCACATCGTTTCCAAGTG 5495
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1658
1639 LeuAlaAlaTyGlnGluGlnLeuAspArgProValValProHisProTyArgVal 1658
QY 5496 GGAGATTCACTACGTAGACGCCACCGTGCAGGAACCTCGAGACTCGGTGGAAGGC 5555
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1678
1659 GlyAspThrValTrpValArgHisGlnThrLysAsnLeuGluProArgTrpLysGly 1678
QY 5556 CTTATCTCTACTTTTGACACACCAACCGCTGTGAAATCGAAGAAATCTCCACTGG 5615
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1698
1679 ProTyThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleAlaAlaTrp 1698
QY 5616 ATCATGTCATCCACGTTAAACGGCGCCACCT-----CCCGATTG-----GGG 5660
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1718
1699 IleHisAlaAlaHisValLysAlaAlaAspProGlyGlyGlyProSerSerArgLeuThr 1718
QY 5661 TGAAGACCCGAAAGACTGAAATCCCTTAAAGCTTCGCTCCATCGCGTGTTCCT 5717
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 1737
1719 TrpArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGluAlaPro 1737

RESULT 3
US-09-718-096-13
; Sequence 13, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718,096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1737
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; OTHER INFORMATION: gag-pol protein
US-09-718-096-13
Alignment Scores:
Pred. No.: 0 Length: 1737
```

```
Score: 5338.00 Matches: 1035
Percent Similarity: 71.6% Conservative: 239
Best Local Similarity: 58.2% Mismatches: 395
Query Match: 36.5% Indels: 110
DB: 2 Gaps: 26

US-10-723-552-3 (1-8132) x US-09-718-096-13 (1-1737)
QY 585 ATGGGACAGCGGTGACGACCCCTTACTTTGACTCTCGACCATTTGGACTGAAGTTAAA 644
DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20
QY 645 TCCAGGGCTCATAAATTTGTTCAGTTTCAGTTTAAGAGGGACCTTTGGCAGACTTTCTGTGTC 704
DB 21 ArgIleAlaHisAsnGlnSerValAspValLysArgArgTrpValThrPheCysSer 40
QY 705 TCTGAATGCGCCGACATTCGATGTTGATGTCGCATCAGAGGGACCTTTAACTTCAGATT 764
DB 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60
QY 765 ATCTCGGCTGTTAAACGAGTTATTTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 824
DB 61 IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80
QY 825 CCTATATCTTACGTGGCAAGATTTTGGCAGAGGATCTCCCGCATCGGTTAAACCATGG 884
DB 81 ProTyIleValThrTrpGluAlaLeuAlaPheAspProProTrpValLysProPhe 100
QY 885 CTGAATTAAGCCAAAGAACGAGTCCCAGAAATTCCTGGCTCTTGGAGAGAAAAACAACAC 944
DB 101 ValHis---ProLysProProProLeuProProSerAlaProSerLeuProLeuGlu 119
QY 945 TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTACCCGAGATTGAGGAGCCA---CCG 1001
DB 120 ProProArgSerThrProProArgSerSerLeuTyProAlaLeuThrProSerLeuGly 139
QY 1002 GCTTGGCCGGAACCCCAA-----GGAGCTCGCGGTGGAGGACCT-----TCT 1022
DB 140 AlaLysProLysProGlnValLeuSerAspSerGlyGlyProLeuIleAspLeuThr 159
QY 1023 GTTCCCGCCACCCCTTATCTGCGACAGGTGCGCGAGGGACCTTTTGCCTCTCT--- 1079
DB 160 GluAspProProProTy-----ArgAspProArgProProProSer 173
QY 1080 -----GGAGCTCGCGGTGGAGGACCT----- 1103
DB 174 AspArgAspGlyAsnGlyGluAlaThrProAlaGlyGluAlaProAspProSerPro 193
QY 1104 --GCTGCGAGGGACTCGGAGCGCGAGGGCGCCACCCCGAGCGGACAGACGATCGCG 1160
DB 194 MetAlaSerArgLeuArgGlyArgArgGluProProValAlaAspSerThrThrSerGln 213
QY 1161 ACATTTACGCTGCGACAGTACGCGCTCCACACCGGGGGGCAATTGACGCCCTCCAG 1220
DB 214 AlaPheProLeuArgAlaGlyGly-----AsnGlyGln-----LeuGln 226
QY 1221 TATTGGCCCTTTTCTCTGAGATCTCTATATTTGGAAGTAACCATCCCTTTCTCTCG 1280
DB 227 TyrTrpProPheSerSerSerAspLeuTyAsnTrpLysAsnAsnAsnProSerPheSer 246
QY 1281 GAGGATCCCCAACGCTCACGGGGTTGGTGGAGTCCCTTATGTTCTCTCACACGACCTACT 1340
DB 247 GluAspProGlyLysLeuThrAlaLeuIleGluSerValLeuIleThrHisGlnProThr 266
QY 1341 TGGGATGATTGTCAAACGCTGTGTCAGACACTCTTCAACACCGAGGAGCGAGAGATT 1400
DB 267 TrpAspAspCysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGluGluLysGlnArgVal 286
QY 1401 CTATTAGCGCTAGAAAAAATGTTCTCGGGCGCGAGCGGCCACCGCGGTGCAAAAT 1460
DB 287 LeuLeuGluAlaArgLysAlaValArgGlyAspAspGlyArgProThrGlnLeuProAsn 306
QY 1461 GAGATTGACATGGGATTTTCCTTAACTCGCCCGGTTGGGACTACAAACACGGCTGAAGGT 1520
```



649	GlyProMetGlyGlnProLeuGlnValLeuThrLeuAsnIleGluAspGluHisArgLeu	668
2547	TACTCTCCCTAGTAAAGCCCTGATCAAAATATA---CAATTCTGGTTGGACAGTTTCCC	2603
669	HisGluThrSerLysGluProAspValSerLeuGlySerThrTrpLeuSerAspPhePro	688
2604	CAAGCCTGGCGAGAAACCGCAGGATGGTGTGGCAAAGCAAGTTCCTCCCAAGTATTATT	2663
689	GlnAlaTrpAlaGluThrGlyGlyMetGlyLeuAlaValArgGlnAlaProLeuIleile	708
2664	CAACTGAAGCCAGTGCACACCACTGTCAGTCAGACAGTACCCCTTGAGTAAGAAGCT	2723
709	ProLeuLysAlaThrSerThrProValSerIleLysGlnTrpMetSerGlnGluAla	728
2724	CAAGAAGCAATTCCGGCGCATGTCCAAAGATTAAATCCACAGGCGCATCTAGTTTCTGTGTC	2783
729	ArgLeuGlyIleLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys	748
2784	CAATCTCCCTGGAACTACCTCCCTGCTACCGGTTAGAAAGCCTGGGACTAATGACTATCGA	2843
749	GlnSerProTrpAsnThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg	768
2844	CCAGTACAGGACTTGAGAGAGGTCATAAACGGGTGCAGGATATACACCCAAAGTCCCG	2903
769	ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro	788
2904	AACCTTATAACCTTGTGTGCTCTCCACCCCAACGGAGCTGGTATACAGATTATGGAC	2963
789	AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp	808
2964	TTAAGGATGCCTTCTTCCTCCCTGAGATTACACCCCACTAGCCAAACACTTTTGGCTTC	3023
809	LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPhe	828
3024	GAATGGAGAGATCCAGTACGGGAGNAACCGGCGAGCTCACCTGGAGCCGACGCCCAA	3083
829	GlnTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln	848
3084	GGGTTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGGCCAACTTC	3143
849	GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe	868
3144	AGGATCCAAACACCTCAGTGCACCTCTCCAGTACGTGGATGACCTGCTTCTGGCGGGA	3203
869	ArgIleGlnHisProAspLeuIleLeuLeuGlnTyr-ValAspAspLeuLeuAlaAla	888
3204	GCCACCAACAGGACTGCTTAGAAGCGCAGCAAGGCACACTACTGCTGGAAATGTCGTACCTA	3263
889	ThrSerGlnLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAsnLeu	908
3264	GGCTACAGACCTCTGCTAAGAAGGCCAGATTTCAGAGAGAGAGGTAAACATACTTGGGG	3323
909	GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly	928
3324	TACAGTTTGGGACGGCGCAGATGGCTGACGGAGGCACGGGAAGAAAACTGTAGTCCAG	3383
929	TyrLeuLeuLysGluGlyGlnArgTyrLeuThrGlnAlaArgLysGluThrValMetGly	948
3384	ATACCGCCGCCAACCAACAGCCAAACAAATGAGAGAGATTTTTGGGGACAGCTGATTTTGC	3443
949	GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys	968
3444	AGACTGTGGATCCCGGGTTTGGACCTTAGCAGCCCCACTCTACCCGGCTAAACCAAGAA	3503
969	ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr	988
3504	AAAGGGGAATTCTCTGGGCTCTGAGCACCAGAGGCAATTTGATGCTATCAAAAAGGCC	3563
989	GlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTyrGlnGluIleLysGlnAla	1008
3564	CTGCTGAGCGCAGCTGCTCTGGCCCTCCCTGACGTAACTAAACCCCTTACCTTTATGTG	3623
1009	LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal	1028

QY	3624	GATGAGCGTAAAGGAGTAGCCCGGGAGT	TTTAAACCCAAACCCCTAGGACCATGGAGAAGA	3683
Db	1029	AspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyP-roTprAsgArg	1048	
QY	3684	CTGTGCGCTACCTGTCAAAAGAGCTCGATCTGTAGCCAGTGGTGGCCCATATGCCTG	3743	
Db	1049	ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTprProCysLeu	1068	
QY	3744	AAGGCTATCGCAGCTGTGGCCATACTGGTCAAGGAGCGCTGACAAATGACTTTGGGACAG	3803	
Db	1069	ArgMetValAlaAlaIleAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGln	1088	
QY	3804	AATATAACTGTAAATAGCCCCCATGCTATGGAGAACATCGTTCGGCAGCCCCCAGACCCGA	3863	
Db	1089	ProLeuValIleLeuAlaProHisAlaValGluAlaLeuValLysGlnProProAspArg	1108	
QY	3864	TGGATGACCAACGCCCGCATGACCCACTATCAAGAGCTGCTTCTC	3920	---ACAGAGAGGCTC
Db	1109	TrpLeuSerAsnAlaArgMetThrHisTyrGlnAlaLeuLeuLeuAspThrAspArgVal	1128	
QY	3921	ACGTTTCGCTCCACGAGCGCTCTCAACCCGCGACTCTTCGTGCTGAAGAGACTGATGAA	3980	
Db	1129	GlnPheGlyProValValAlaLeuLeuProIleAlaLeuLeuPro	1147	---LeuProGluGlu
QY	3981	CCAGTGACTCATGATTGCCATCAACTATGATTGAGGAGACTGGGTCCGCAAGGACCTT	4040	
Db	1148	GlyLeuGlnHisAsnCysLeuAspIleLeuAlaGluAlaHisGlyThrArgProAspLeu	1167	
QY	4041	ACGACATACCGCTGACTGCGAGAGTGCTAACTGGTTCCTACGCGAAGACGCTATGTG	4100	
Db	1168	ThrAspGlnProLeuProAspAlaAspHisThrTprTyrThrAspGlySerSerLeuLeu	1187	---:::
QY	4101	GTGGAAGGTAAAGAGTAGGCTGGCGCGCGGTGGTGGACGGGACCCGACGATCTGGGCC	4160	
Db	1188	GlnGluGlyGlnArgLysAlaGlyAlaAlaValThrThrGluThrGluValIleTprAla	1207	
QY	4161	AGCAGCTCGCGGAAGAACTTCAGACAACAAAGCTGAGTCTATGGCCCTCAGCGAAGCT	4220	
Db	1208	LysAlaLeuProAlaGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAla	1227	
QY	4221	TTGCGGCTGCGCGAAGGAAATCCATAAACATTTATACGACAGCAGCGATGCTTTCGG	4280	
Db	1228	LeuMetAlaGluGlyLysLysLeuAsnValTyrThrAspSerArgTyrAlaPheAla	1247	
QY	4281	ACTGCACAGCTACATGGGGCCATCTATAACAAAGGGGGTTCCTTACCTCAGCAGGGAGG	4340	
Db	1248	ThrAlaHisIleHisGlyGluIleTyrArgArgGlyLeuLeuThrSerGluGlyLys	1267	:::~::~:
QY	4341	GAATATAAGAACAAAGAGGAATCTTAAGCTATTAGACCGGTACATTTACCAAAAGG	4400	
Db	1268	GluIleLysAsnLysAspGluIleLeuAlaLeuLeuLysAlaLeuPheLeuProLysArg	1287	~::~:~::~:
QY	4401	CTAGCTATTATACATCTCTCGACATCAGAAAGCTCTCATATCCAGGGAAC	4460	
Db	1288	LeuSerIleIleHisCysProGlyHisIleGlnLysGlyHisSerAlaGluAlaArgGlyAsn	1307	~::~:~::~:
QY	4461	CAGATGCTGACCGGGTTGCCAAGCAGGACGCCAGGGGTGTTAACCTTCTGCCTATAATA	4520	~::~:~::~:
Db	1308	ArgMetAlaAspGlnAlaAlaArgLysAlaAla	1320	~::~:~::~:~::~:~::~:IleThr
QY	4521	GAATGCCCAAGGCC	4547	-----CCAGAACCCAGA
Db	1321	GluThrProAspThrSerThrLeuLeuIleGluAsnSerSerProTyrThrSerGluHis	1340	~::~:~::~:
QY	4548	CGACAGTACACCTTAGAAGACTGCGAAGAGATAAAAGATAGACACAG	4604	---TTCTCTGAG
Db	1341	PheHisTyrThrValThrAspIleLysAspLeuThrLysLeuGlyAlaIleTyrAspLys	1360	~::~:~::~:~::~:~::~:
QY	4605	ACTCGGAAGGACCTGCTATACCTCAGATGGGAGGAATCTCGCCCCCAAGAAGGG	4664	
Db	1361	ThrLysLysTyrTprValTyr-----GlnGlyLysProValMetProAspGlnPheThr	1378	~::~:~::~:~::~:~::~:

US-09-075-272-4  
; Sequence 4, Application US/09075272  
; Patent No. 6136598  
; GENERAL INFORMATION:  
; APPLICANT: MILLER, A. DUSTY  
; APPLICANT: WOLGAMOT, GREG  
; APPLICANT: BOWHAM, LYNN  
; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL  
; TITLE OF INVENTION: PACKAGING CELL LINES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/075,272  
; FILING DATE: 08-MAY-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/046,140  
; FILING DATE: 09-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Poor, Brian W.  
; REGISTRATION NUMBER: 32,928  
; REFERENCE/DOCKET NUMBER: 14538A-003710  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 467-9600  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1203 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-075-272-4

Alignment Scores:  
Pred. No.: 0 Length: 1203  
Score: 4329.50 Matches: 822  
Percent Similarity: 79.6% Conservative: 129  
Best Local Similarity: 68.8% Mismatches: 222  
Query Match: 29.6% Indels: 21  
DB: 2 Gaps: 7

US-10-723-552-3 (1-8132) x US-09-075-272-4 (1-1203)

QY 2160 GGGAGACGGGTTTCGACCCCTCCCGAGCCCGGTAACCTTTGAAGTGGAGGGGCAA 2219  
Db 1 GlySerArgGlySerAspProLeuProGluProArgValThrLeuSerValGluGlyThr 20  
QY 2220 CCAGTTGACCTCTGCTGATACCGGAGCGAAACATTGCTGCTACTACAGCCATTAGGA 2279  
Db 21 ProValAsnPheLeuLeuAspThrGlyAlaGluHisSerValLeuThrSerProLeuGly 40  
QY 2280 AAACTAAAAGATAAAAATCCCTGGGTGATGGGTGCGACAGGGCAACACAGTATCCATGG 2339  
Db 41 LysLeuGlySerLysArgThrIleValValGlyAlaThrGlySerLysLeuTyrProTyr 60  
QY 2340 ACTACCCGAGAACAGTTGACCTGAGTGGGAGTGGGACCGGTAACCCACTCGTTCTGCTGTCATA 2399  
Db 61 ThrThrLysArgAlaLeuGlnIleAspLysAsnMetValThrHisSerPheLeuValIle 80  
QY 2400 CCTGAGTGGCCAGCACCCCTCTTAGGTAGACAGATTATGACCAAGATGGGACACAAATT 2459  
Db 81 ProGluCysProAlaProLeuLeuGlyArgAspLeuLeuThrLysLeuLysAlaGlnVal 100

QY 2460 TCTTTTGAACAGGGAAACACAGAAAGTGTCTGCAATAATAACAAACCTATCACTGTGTGACC 2519  
Db 101 GlnPheThrSerGluGlyProGlnValSerTrpGlyLysAlaProLeuAlaCysLeuVal 120  
QY 2520 CTCCAATTAGATGACGAATATCGACTATCTCTCCCTAGTAAAGCTGATCAAAATATA 2579  
Db 121 LeuSerThrGluGluGlyTrpArgLeuHisGluGluGlnProLysGlyAlaAlaProLeu 140  
QY 2580 CAATTCTGGTGGAAACAGTTTCCCAAGCTGGCGAGAAACCGCAGGATGGGTTGGCA 2639  
Db 141 Asp---TrpValThrAlaPheProAsnValTrpAlaGluGlnAlaGlyMetGlyLeuAla 159  
QY 2640 AAGCAAGTTCCCCCAACAGTTATTCACTGAAGCCAGTGCCACACAGTGTCTCAGTCAGA 2699  
Db 160 LysGlnValProProValValValGluLeuLysAlaAspAlaThrProLysSerValArg 179  
QY 2700 CAGTACCCCTTGGTAAAGAGCTCAAGAGGAATTCGGCCGCATGTCCTCAAGATTAATC 2759  
Db 180 GlnTyrProMetSerLysGluAlaLysGluGlyIleArgProHisIleArgArgLeuLeu 199  
QY 2760 CAACAGGGCATCTAGTCTCTGTCCTCAATCTCCCTGGAATCTCCCTGCTACCGGTTAGA 2819  
Db 200 AspGlnGlyIleLeuValAlaCysGlnSerProTrpAsnThrProLeuLeuProValArg 219  
QY 2820 AAGCCTGGGACTAATGACTATCGACAGTACAGGACTTCAGAGAGGTCAATAAACCGGTG 2879  
Db 220 LysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsnLysArgVal 239  
QY 2880 CAGGATATACACCAACAGTCCCGAACCTTATACTCTTGTGTCTCTCCACCCCAA 2939  
Db 240 LeuAspIleHisProThrValProAsnProTyrAsnLeuLeuSerSerLeuProProGlu 259  
QY 2940 CGAGCTGTATACAGATTGGACTTAAAGGATGCTCTCTCTGCTGAGATTTACACCCC 2999  
Db 260 ArgThrTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisPro 279  
QY 3000 ACTAGCCACCACTTTTTCCTTCCGAATCGAGAGATCCAGGTACGGGAAGAACCGGGCAG 3059  
Db 280 LysSerGlnLeuLeuPheAlaPheGluTrpArgProGluGlyGlnThrGlyGln 299  
QY 3060 CTCACCTGGACCCGACTGCCCCAAGGTTCAAGAACTCCCGACCACTTTTTCAGCAAGCC 3119  
Db 300 LeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAspGluAla 319  
QY 3120 CTACACAGAGACTGCGCAACTTCAGGATCCACACCTCAGTGCACCTCTCCAGTAC 3179  
Db 320 LeuHisArgAspLeuAlaProPheArgAlaGlnAsnProGlnLeuThrLeuLeuGlnTyr 339  
QY 3180 GTGGATGACCTCTCTGCGGGAGCCACAAACAGGACTGCTTAGAAGGCACGAAGCA 3239  
Db 340 ValAspAspLeuLeuIleAlaAlaSerLysGluLeuCysGlnGlnGlyThrGluArg 359  
QY 3240 CTACTGCTGGAATGCTGACCTAGCTACAGAGCTCTGCTAAAGAGCCCAAGATTTGC 3299  
Db 360 LeuLeuThrGluLeuGlyAsnLeuGlyTyrArgValSerAlaLysLysAlaGlnIleCys 379  
QY 3300 AGGAGAGGTAACATCTTGGGTGACAGTTTGGGAGCGGCGGACGATGGCTGACGGAG 3359  
Db 380 GlnThrGluValIleTyrLeuGlyTyrThrLeuArgGlyLysGlyLysArgTrpLeuThrGlu 399  
QY 3360 GCACGGAAGAAAACCTGTAGTCCAGATACCGGGCCCAACACACAGCCAAACAAATGAGAGAG 3419  
Db 400 AlaArgLysLysThrValMetMetIleProProProThrThrProArgGlnValArgGlu 419  
QY 3420 TTTTGGGACAGCTGGATTTTTCAGACTGTGGATCCCGGGGTTTTCGACCTTAGCAGCC 3479  
Db 420 PheLeuGlyThrAlaGlyPheCysArgLeuTyrIleProGlyPheAlaThrLeuAlaAla 439  
QY 3480 CCACCTCTACCCGCTAACCAAGAAAAGGGGAATTTCTCTGGGCTCTCTGAGCAGCAGAG 3539  
Db 440 ProLeuTyrProLeuThrArgGluGlyIleProPheGluTrpLysGluGluHisGlnArg 459





```
QY 3951 GCCACTCTCTCTGCTGAAGAGACTGATGAACCACTGACTCATGATTGGCCATCAACTATTG 4010
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
481 AlaThrLeuLeuPro---LeuProGluGluGluGluGlnHisAsnCysLeuAspIleLeu 499
QY 4011 ATTGAGGAGACTGGGGTCCCAAGGACTTTACAGACATACATCCGCTGACTGGAGAGTGCTA 4070
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
500 AlaGluAlaHisGlyThrArgProAspLeuThrAspGlnProLeuProAspAlaAspHis 519
QY 4071 ACCTGGTTCACTACGGAACAGCTATGCTGGTGAAGGTAAAGGATGGCTGGGGGGCGG 4130
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
520 ThrTrpTrpThrAspGlySerLeuLeuGlnGluGlyGlnArgLysAlaGlyAlaAla 539
QY 4131 GTGGTGAACGGGACCCGACAGATCTGGGCCAGCAGCTCGCCGGAAGAACTTCAGACACA 4190
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
540 ValThrThrGluThrGluValIleTrpAlaLeuAlaLeuProAlaGlyThrSerAlaGln 559
QY 4191 AAGCTGAGTCTATGGCCCTCAGCAAGCTTTGGGCTGCCGCAAGGAAATCCATAAAC 4250
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
560 ArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLysLysLeuAsn 579
QY 4251 ATTATACGACAGCAGGATGCTTTGGCTGCGACTGCACAGTACATGGGCCATCTATAA 4310
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
580 ValTrpThrAspSerArgTrpAlaPheAlaThrAlaHisIleHisGlyGluIleTrpArg 599
QY 4311 CAAAGGGGTGTCTTACCTCAGCAGGGAGGAAATAAAGAACAAAGAGGAAATCTTAAGC 4370
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
600 ArgArgGlyLeuLeuThrSerGluGlyLysGluIleLysAsnLysAspGluIleLeuAla 619
QY 4371 CTATTAGAAGCCGTACATTACCAAAAGGCTAGCTATTATACATGCTCTCGACATCAG 4430
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
620 LeuLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysProGlyHisGln 639
QY 4431 AAGCTAAAGATCTATATCCAGAGAAACAGATGCTGACCGGTGGCCAGCAGGCA 4490
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 LysGlyHisSerAlaGluAlaArgGlyAsnArgMetAlaAspGlnAlaAlaArgLysAla 659
QY 4491 GCCAGGGTGTAACTTCTGCTCTAATAAGAAATGCCAAAGCC----- 4535
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
660 Ala-----IleThrGluThrProAspThrSerThrLeuLeuIle 672
QY 4536 -----CCAGAACCCAGACGACGATACACCTAGAACAGCTGGCAAGAG 4577
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
673 GluAsnSerSerProTrpThrSerGluHisPheHisTrpValThrAspIleLysAsp 692
QY 4578 ATAAAAGATAGACCAAG-----TTCTCTGACACTCCGGAAGGACCTCTATACCTCAGAT 4634
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
693 LeuThrLysLeuGlyAlaIleTrpAspLysThrLysLysTrpValTyr-----Gln 710
QY 4635 GGAAGGAAATCTGCCCCCAAGAGGGTTAGAAATATGTCACACAGATACATCGTCTA 4694
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
711 GlyLysProValMetProAspGlnPheThrPheGluLeuAspPheLeuHisGlnLeu 730
QY 4695 ACCCACTTAGGAACCTAACACCTGCAGCAGTTGTCAGAAC-----TCCCTTTATCAT 4748
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
731 ThrHisLeuSerPheSerLysMetLysAlaLeuLeuGluArgSerHisSerProTrpTyr 750
QY 4749 GTTCTGAGGTACACAGAGTGGCTGACTCGGTGTGTCAAACATTTGTGTGCCCTGCCAGCTG 4808
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
751 MetLeuAsnArgAspArgThrLeuLysAsnIleThrGluThrCysLysAlaCysAlaGln 770
QY 4809 GTTAATGCTAATCTCTCCAGAACTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGGC 4868
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
771 ValAsnAlaSerLysSerAlaValLysGlnGlyThrArgValArgGlyHisArgProGly 790
QY 4869 GCTCACTGGGAAGTGGACTTCACTGAGTAAAGCCGCTAAATACGGAACAAATACCTA 4928
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
791 ThrHisTrpGluIleAspPheThrGluIleLysProGlyLeuTrpGlyTyrLysTyrLeu 810
QY 4929 TTGGTTTTTGTAGACACTTTTCAGGATGGGTAGAGCTTATCTCTACTAAGAAAGACT 4988
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
811 LeuValPheIleAspThrPheSerGlyTrpIleGluAlaPheProThrLysLysGluThr 830
QY 4989 TCAACCGTGTGGCTAAAAAATACTGGAAGAAATTTTCCAGATTTTGGAAATACCTAAG 5048
```

```
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
831 AlaLysValValThrLysLysLeuLeuGluIlePheProArgPheGlyMetProGln 850
QY 5049 GTAATAGGTGACACAATGCTCAGCTTTTGTGGCCAGGTAAGTCAGGACTGGCCAAG 5108
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
851 ValLeuGlyThrAspAsnGlyProAlaPheValSerLysValSerGlnThrValAlaAsp 870
QY 5109 ATATTGGGATTCATTGGAACTGCTATTGTCATACAGACCCCAAGCTCAGACAGGTA 5168
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
871 LeuLeuGlyIleAspTrpLysLeuHisCysAlaTrpArgProGlnSerGlyGlnVal 890
QY 5169 GAGAGATGAATAGAACCACTTAAAGAGACCTTTACTATAATTGACCCGGAGACTGCGTT 5228
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
891 GluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThrLeuAlaThrGlySer 910
QY 5229 AATGATTGATAGTCTCTCTGCTCTGCTTTTGGTGTAGGTAACACCCCTGGACAG 5288
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
911 ArgAspTrpValLeuLeuLeuProLeuAlaLeuTrpArgAlaArgAsnThrProGlyPro 930
QY 5289 TTTGGGCTGACCCCTATCAATTTACTCTACGGGGGACCCCTTGTCTCTAGGCTCAAGCACTT 5348
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
931 HisGlyLeuThrProTrpGluIleLeuTrpGlyAlaProProLeuValAsnPhePro 950
QY 5349 TCTGTACATAGTGTGCTGCTGCTTCTCCAGCCTTTGTTCTCTAGGCTCAAGCACTT 5408
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
951 AspProAspMetThrArgValThrAsnSerProSerLeuGlnAlaHisLeuGlnAlaLeu 970
QY 5409 GAGTGGGTGAGACAACAGCGTGGAGGCACTCCGGGAGGCTCTACTCAGGAGGAGAGAC 5468
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
971 TyrLeuValGlnHisGluValTrpArgProLeuAlaAlaIleTrpGlnGluLeuAsp 990
QY 5469 TTGCAG--ATCCCATCATGTTTCCAAAGTGGGAGATTCTAGCTTACGTAGACGCCACCGT 5525
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
991 ArgProValValProHisProTrpArgValGlyAspThrValTrpValArgArgHisGln 1010
QY 5526 GCAGGAAACCTCGAGACTCGGTGGAGGCGCTTATCTCGTACTTTTGACCACCAACAG 5585
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1011 ThrLysAsnLeuGluProArgTrpLysGlyProTyrThrValLeuLeuThrThrProThr 1030
QY 5586 GCTGTGAAGTGAAGAAATCTCCACTGATCCATCCATCCATCCATCCATCCATCCATCCATCC 5645
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1031 AlaLeuLysValAspGlyIleAlaAlaTrpIleHisAlaAlaHisValLysAlaAlaAsp 1050
QY 5646 CCT-----CCCGATTG-----GGGTGAAAGCCGMAAAGACTGAAAATCCCTT 5690
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1051 ProGlyGlyGlyProSerSerArgLeuThrTrpArgValGlnArgSerGlnAsnProLeu 1070
QY 5691 AAGCTTTCGCTCCATCCGCTGCTTCT 5717
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1071 LysIleArgLeuThrArgGluAlaPro 1079

RESULT 6
US-09-376-784-6
; Sequence 6, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION
; APPLICANT: Benerio, Papi T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 638
; TYPE: PRN
; ORGANISM: Artificial Sequence
```

FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PERV-C  
; OTHER INFORMATION: polypeptide sequence taken from GenBank Accession  
; OTHER INFORMATION: No. 6261806 AF038600 for comparison.  
US-09-376-781-6

Alignment Scores:  
Pred. No.: 8.95e-285 Length: 638  
Score: 3425.00 Matches: 637  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.8% Mismatches: 0  
Query Match: 23.4% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3 (1-8132) x US-09-376-781-6 (1-638)

QY 5620 ATGCATCCCAAGTTAAACCGCGCCACCTCCCGATTCGGGTGGAAGCGGAAGACACTG 5679  
DB 1 MetHisProThrLeuAsnArgHisLeuProIleArgGlyLysProIleArgLeu 20  
QY 5680 AAAATCCCTTAAGCTTCATCGCTCCATCGCTGGTTCCTTACTCTGTCAATAAACCCTCTCAG 5739  
DB 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
QY 5740 ACTAATGGTATGGCATAGGAGACAGCTGAACTCCCATAAACCCTTATCTCAGCTGG 5799  
DB 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 5800 TTAATTACTACCTCCGACAGGTATTAAATATCAACAACACTCAAGGGAGGCTCCTTTA 5859  
DB 61 LeuIleThrAspSerGlyThrGlyLeAsnIleAsnAsnThrGlnGlyAlaProLeu 80  
QY 5860 GGAACCTGGTGGCTGATCTATACGTTCCTCAGATCAGTTATTCCTAGTCTGACCTCA 5919  
DB 81 GlyThrTrpProAspLeuValCysLeuArgSerValIleProSerLeuThrSer 100  
QY 5920 CCCCAGATATCTCATGCTCAGGATTTATTTGTTGGCCAGGACCAACAAATATGGA 5979  
DB 101 ProProAspIleLeuHisAlaHisGlyPheTyValCysProGlyProProAsnAsnGly 120  
QY 5980 AAACATTGGGAATCCAGAGATTTCTTTTGAACAACTGGAACCTGTAACCTCTAAT 6039  
DB 121 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 140  
QY 6040 GATGGATATTGGAATGGCCAACTCTCAGCAGGATAGGTAAAGTTTCTTATGTC AAC 6099  
DB 141 AspGlyTyTrpLysTrpProThrSerGlnAspArgValSerPheSerTyValAsn 160  
QY 6100 ACCTATACAGCTCTGGACAATTAATTACCTGACCTGGATTAGACTGGAAGCCCAAG 6159  
DB 161 ThrTyThrSerSerGlyGlnPheAsnTyLeuThrTrpIleArgThrGlySerProLys 180  
QY 6160 TGCTCTCCTCAGACTAGATTACCTAAATAAGTTTCACTGAGAAGGGAACAGAA 6219  
DB 181 CysSerProSerAspLeuAspTyLeuLysIleSerPheThrGlnLysGlyLysGlnGlu 200  
QY 6220 AATATCTTAAATGGGTAAATGGTATGCTTGGGGAATGGTATATTATGAGGCTCGGT 6279  
DB 201 AsnIleLeuTyTrpValAsnGlyMetSerTrpGlyMetValTyTyTrpGlySerGly 220  
QY 6280 AAACAACAGGCTCCATTTCACTATTTCGCTCAAAATAAACAGCTGAGCTCCAAATG 6339  
DB 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProMet 240  
QY 6340 GCTATAGGCCAATAACGCTTCGCGGTCAAGACCCCAACCCCAAGGACCCAGACCA 6399  
DB 241 AlaIleGlyProAsnThrValLeuThrGlnArgProProProProGlyProGlyPro 260  
QY 6400 TCCTCTAAACATAACTTCTGGATCAGACCCCACTGAGTCTAAACAGACCACTAAATGGGG 6459  
DB 261 SerSerAsnIleThrSerGlySerAspProThrGluSerSerSerThrThrLysMetGly 280  
QY 6460 GCAAAACTTTTAGCTCATCCAGGGAGCTTTTCAAGCTCTTAACCTCCAGACTCCAGAG 6519

RESULT 7

US-09-111-085-2

DB 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 6520 GCTACCTCTTCTTGGTCTATGCTTACCTTCCGGGCCACCTTACTATGAAGAAATGGCT 6579  
DB 301 AlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyTyTrpGluGlyMetAla 320  
QY 6580 AGAAGAGGGAATTCATGTGCAAAAGAACATAGAGACCAATGCACATGGGATCCCAA 6639  
DB 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 6640 AATAAGCTTACCTTACTGAGGTTCTGGAAAAGGACCTGCATAGAAAAGTTCCCCCA 6699  
DB 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 6700 TCCACCAACACACTTGTAAACACACTGAAGCCTTTAATCAACACTCTGAGAGTCAATAT 6759  
DB 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTy 380  
QY 6760 CTGTACTCTGGTTATGACAGGTGGTCATGTAATCTAGGATTAACCCCTTGTGTTTCC 6819  
DB 381 LeuValProGlyTyTrpAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 6820 ACCTTGGTTTAAACCAAACTAAAGATTTTTCATTTATGGTCCAAATTTGTTCCCCAGTG 6879  
DB 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 6880 TATTACTATCCCGAAAACCAATCCTTGATGATATGACTACAGAAATCATCCACAAAAG 6939  
DB 421 TyTyTyTyProGluLysAlaIleLeuAspGluTyAspTyArgAsnHisArgGlnLys 440  
QY 6940 AGAAGACCCATATCTCTGACACTGTCTGTGATGCTCGACTTGGAGTGGCAGAGGTGA 6999  
DB 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaGlyVal 460  
QY 7000 GGAACAGGAACAGCTGCCCTGTCCAGGACCAACAGCAGCTAGAAAACAGACTTAGTAAC 7059  
DB 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 7060 CTACATCGAATTTGAACAGAGATCTCAAGCCCTAGAAAATCTCTCAGTACCTGGAG 7119  
DB 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 7120 GAATCCCTAACTCTCTTATCTGAAGTAGTCTTACAGATAGAGAGGGTTAGATTATTA 7179  
DB 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
QY 7180 TTTCTAAAGAGAGGAGATTATGTGTAGCTTGAAGGAGAAATGCTGTTTTATGTGGAT 7239  
DB 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyValAsp 540  
QY 7240 CATTGAGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAGGTTGGAGAGCCCTCGA 7299  
DB 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgTy 560  
QY 7300 AGGAAAAGAAAACCTACTCAAGGCTGTTTGAAGGATGTTCAACAGGCTCTCTTTGGTTG 7359  
DB 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeu 580  
QY 7360 GCTACCTTACTTCTGCTTTTAAACAGGACCTTAATAGTCTCTCTCTCTGTTTACTCAGTT 7419  
DB 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
QY 7420 GGGCCATGTATTATTAAACAAGTTAATTCCTTTCATTAGAGAAGTAATGTCAGTCCAG 7479  
DB 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 7480 ATCAGGTACTTAGACAACAGTACCAAGCCGCTTAGCAGGGAAGCTGGCGC 7533  
DB 621 IleMetValLeuArgGlnGlnTyGlnSerProSerSerArgGluAlaGlyArg 638



```
; Sequence 2, Application US/091111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Porcine retrovirus
US-09-111-085-2

Alignment Scores:
Pred. No.:      8.18e-237      Length:      660
Score:          2866.50      Matches:      537
Percent Similarity: 88.6%      Conservative: 39
Best Local Similarity: 82.6%      Mismatches: 53
Query Match:     19.6%      Indels:      21
DB:              2          Gaps:         6

US-10-723-552-3 (1-8132) x US-09-111-085-2 (1-660)
QY 5620 ATGCATCCCAAGTAAACCGCGCCACCTCCGATTCGGGGTGGAAAGCCGAAAGACTG 5679
DB 1 MetHisProThrLeuSerArgArgHisLeuProIleArgGlyGlyProLysArgLeu 20
QY 5680 AAAATCCCTTAAGCTTCGCTCCATCGCTGCTTCTTACTCTGTCGAATAACCTCTCAG 5739
DB 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
QY 5740 ACTAATGGTATGGCATAGAGACAGCCTGAACCTCCCATAAACCTTATCTCTCACCTGG 5799
DB 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
QY 5800 TTAATTAAGTACTCCGCGCAGAGTATTAATATCAACAACACTCAAGGGGAGGCTCCCTTTA 5859
DB 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyAlaProLeu 80
QY 5860 GGAACCTGGTGGCTGATCATACGTTTGCTCGCTCAGATCAGTTATCTAGTCTG----- 5913
DB 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY 5914 ----ACCTCACCCCGAGTATCCCTCCATGCTCAGGATTTTATGTTGCCCAGGACCA 5970
DB 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
QY 5971 AATAATGGAAACATTCGCGAAATCCAGAGATTCTTTTGTAAACAATGGAACTGTGTA 6030
DB 121 AsnAsnGluLeuTyrCysGlyAsnProGlnAspPheCysLysGlnTrpSerCysIle 140
QY 6031 ACCTCTAATGATGATATTCGAAATGGCCAACTCTCAGCAGGATAGGGTAAGTTTCT 6090
DB 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160
QY 6091 TATGTCACACCTATACCACTCTGGACATTTAATTAC-----CTGACC 6135
DB 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
QY 6136 TGG-----ATTAGAACTGGAAAGCCCAAGTGCTCTCTCTTCAGAC 6174
DB 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
QY 6175 CTAGATTACATAAAATGTTTCTACGTAGAGAAAGGAAACAAGAAATATCTCTAAATGG 6234
DB 201 LeuAspTyrLeuLysIleSerPheThrThrGluLysGlyLysGlnGluAsnIleGlnLysTrp 220
```

```
QY 6235 GTAAATGGTATGCTTTGGGAAATGGTATATTATGGAGGCTCGGGTAAACAACACGAGCTCC 6294
DB ValAsnGlyIleSerTrpGlyIleValTyrTyrGlySerGlyArgLysLysGlySer 240
QY 6295 ATTCTAACTATTGCGCTCAAAATA----AACAGCTGGAGCCTCCAATGGCTATAGGACCA 6351
DB ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProValAlaIleGlyPro 260
QY 6352 AATACGGTCTTGACGGGTCAAGACCCCAACCCAA-----GGACGAGGACCA 6399
DB AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGluArgProSerProAsnPro 280
QY 6400 TCCTCT--AACATACTTCTGGATCAGACCCCACTAGTCTAACAGCAGCAGCTATAAATG 6456
DB SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300
QY 6457 GGGGCAAAATTTTATAGCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACGACTCCA 6516
DB GlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320
QY 6517 GAGGCTACCTCTTCTTGGTATGCTTAGCTTCGGGCCCACTTACTATGAGAATG 6576
DB GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMet 340
QY 6577 GCTAGAGAGGGAAATTCATATGTGACAAAGAACATAGAGACCAATGCACATGGGATCC 6636
DB AlaArgGlyGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySer 360
QY 6637 CAAAATAAGCTTACCCCTTACTGAGGTTCTGGAAGAGCCACCTGCATAGAGAAAGTTCCC 6696
DB GlnAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyMetValPro 380
QY 6697 CATTCCCAACCAACACTTTGTAAACCACTGAAGCCCTTTAATCAAACTCTGAGACTCA 6756
DB ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400
QY 6757 TATCTGTAACCTGCTTATGACAGTGGTGGGCAATGTAATACTTGATTAACCCCTTGTT 6816
DB TyrLeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVal 420
QY 6817 TCACCTCTGGTTTTTAAACCAACTAAAGATTTTTGCATTATGTCCTCAAAATTTGTTCCCGA 6876
DB SerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArg 440
QY 6877 GTGTATTACTATCCGAAAAAGCAATCTTGTATGATATATGACTACAGAAATCATCCACAA 6936
DB ValTyrTyrTyrProGluLysAlaValLeuAspGluTyrAspTyrArgTyrAsnArgPro 460
QY 6937 AAGAGAGAACCCATATCTCTGACACTTGTGTGATGCTCGGACTTGGAGTGGCAGCAGGT 6996
DB LysArgGluProLysSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIleGly 480
QY 6997 GTAGGAACAGGACACCTGCTGCTGTCACGGGACCAACAGCAGCTAGAAAACAGGACTTAGT 7056
DB ValGlyThrGlyThrAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSer 500
QY 7057 AACCTACATGAAATTTAAGAGATCTCCAAAGCCCTAGAAAATCTGTCAGTAACCTG 7116
DB AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520
QY 7117 GAGGAATCCCTAACCTCTTATCTGAAGTAGTCTCAGAAATAGAGAGGTTAGATTTA 7176
DB GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 540
QY 7177 TTATTTCTAAAGAGGAGGATTTATGTAGCCCTCAAGAGGAATGCTGTTTTTATGTG 7236
DB LeuPheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrVal 560
QY 7237 GATCATTCAGGGCCATCAGAGACTCCATGACAAGCTTAGAAAAGTTGGAGACGCT 7296
DB AphHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArgArg 580
```



QY 6937 AAGAGAGAACCCATATCTCTGACACTTGTGTGATGCTCGACTTGGAGTGGCAGCAGGT 6996  
DB 461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGly 480  
QY 6997 GTAGGAACAGAACAGCTGCCCTGGTCA CGGACACAGAGCTAGAAA CAGGACTTAGT 7056  
DB 481 ValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSer 500  
QY 7057 AACCTACATCGAATGTAAACAGAGAGTCTCCAGACCCCTAGAAAATCTGTCACTAACCTG 7116  
DB 501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520  
QY 7117 GAGGAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGATAGAGAGGTTTAGATTTA 7176  
DB 521 GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgGlyLeuAspLeu 540  
QY 7177 TTATTTCTAAAGAGAGGATTATGTGTAGCTTGAAGGAGAAATGCTGTTTTATGTG 7236  
DB 541 LeuPheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysPheTyrVal 560  
QY 7237 GATCATTACGGGCCCATCAGAGACTCCATGAA CAAGCTTAGAAAAGGTTGGAGAACGCT 7296  
DB 561 AspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArg 580  
QY 7297 CGAAGGGAAGAAAGAACTACTCAAGGTGGTTTGAGGGATGGTTCAACAGGTCTCTTTGG 7356  
DB 581 ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrp 600  
QY 7357 TTGGCTACCCCTACTTTCTGTTTAAACAGGACCCCTTAATAGTCTCTCTGTTTACTCACA 7416  
DB 601 MetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuThr 620  
QY 7417 GTTGGCCGATGATTATTACAAGTTAAATGTCCTTCAATAGAGAACGAATAAGTGCAGTC 7476  
DB 621 ValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaVal 640  
QY 7477 CAGATCATGTACTTACACAAACAGTACCAA 7506  
DB 641 GlnIleMetValLeuArgGlnGlnTyrGln 650

RESULT 9  
US-09-171-553B-4  
; Sequence 4, Application US/09171553B  
; Patent No. 6756227  
; GENERAL INFORMATION:  
; APPLICANT: GALBRAITH, DANIEL N.  
; APPLICANT: HAWORTH, CHRISTINE  
; APPLICANT: LEES, GILLIAN M.  
; APPLICANT: SMITH, KENNETH T.  
; TITLE OF INVENTION: PORCINE RETROVIRUS  
; FILE REFERENCE: CFPV-5.01  
; CURRENT APPLICATION NUMBER: US/09/171,553B  
; CURRENT FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: PCT/GB97/01087  
; PRIOR FILING DATE: 1997-04-18  
; PRIOR APPLICATION NUMBER: GB 9702668.6  
; PRIOR FILING DATE: 1997-02-10  
; PRIOR APPLICATION NUMBER: GB 9608164.1  
; PRIOR FILING DATE: 1996-04-19  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Porcine retrovirus  
US-09-171-553B-4

Alignment Scores:  
Pred. No.: 1,97e-217 Length: 524  
Score: 2640.00 Matches: 495  
Percent Similarity: 97.3% Conservative: 16  
Best Local Similarity: 94.3% Mismatches: 13  
Query Match: 18.0% Indels: 2

DB: 2 Gaps: 0  
US-10-723-552-3 (1-8132) x US-09-171-553B-4 (1-524)  
QY 585 ATGGGACAGACGGTGTAGCAGCCCTCTTAGTTTGTACTCTCGACCATTTGGACTGAAGTTAAA 644  
DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20  
QY 645 TCAGAGGCTCATAAATTTGTTCAGTTTCAGTTTAAAGAGGACCTTGGCAGACTTTCTGTGTC 704  
DB 21 SerArgAlaHisAsnLeuSerValGlnValLysGlyProTrpGlnThrPheCysAla 40  
QY 705 TCTGAATGCCCCACATTCGATGTCGATGCGCATCAGAGGGGACCTTTAATCTCGAGATT 764  
DB 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
QY 765 ATCTCGCTGTAAACAGCTTATTTTTCAGACTGTGACCCCGCTCTCATCCCGATCAGGAG 824  
DB 61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
QY 825 CCTTATATCTTACGTGGCAAGATTTGGCAGAGGATCCTCCGCCATGGGTTAAACCATGG 884  
DB 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100  
QY 885 CTGAATAAGCAAGAAAGCCAGGTCCTCCGAAATTCCTGCTCTTGGAGAGAAAAACAAACAC 944  
DB 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
QY 945 TCGGCTGAAAAAGTCAAGCCCTC - TCCTCATATCTACCCCGAGATTGAGGAGCCACCGGC 1003  
DB 121 SerAlaGluLysValGluProSerSerTyrLeuProArgAspArgGlyAlaAlaAsp 140  
QY 1004 TTGGCCGGAACCCCAATCTGTTCCCGACCCCTTATCTGGCACAGGTCGCCCGGGGG 1063  
DB 141 LeuAlaGlyThrProThrCysSerProThrProLeuSerSerThrGlyCysGly 160  
QY 1064 ACCCTTTGCCCTCTCTGGAGCTCCGGCGGTGGAGGACCTGCTGCGAGGACTCGAGGCCG 1123  
DB 161 Thr-SerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerAr 180  
QY 1124 GAGGGCGCCACCCCGGAGCGGACAGACGAGATTCGCGACATTACCGCTGGCGACGTACGG 1183  
DB 180 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGl 200  
QY 1184 CCTCCCAACACCGGGGGCCAAATTCAGCCCTCCAGTAGTTGGCCCTTTCTCTCAG 1243  
DB 200 yProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAs 220  
QY 1244 TCTCTATAATTGGAAAACTAACCATCCCTCTTCTCGAGGATCCCAACGCTCACGGG 1303  
DB 220 pLeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGl 240  
QY 1304 GTTGTGGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGATGATGTCAACAGCTGCT 1363  
DB 240 yLeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLe 260  
QY 1364 GCAGACACTCTTCAACACCGAGGAGGAGAGAAATCTATTAGAGGCTAGAAAAATGT 1423  
DB 260 uGlnThrLeuPheThrThrGluGluArgGluIleLeuLeuGluAlaArgLysAsnVa 280  
QY 1424 TCCTGGGGCGGACGGCGACCCACGCGGTTGAAAAATGAGATTGACATGGGATTTCCTT 1483  
DB 280 lProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLe 300  
QY 1484 AACTCGCCCGGTTGGGACTTACAACACCGCTGAAGGTAGGAGAGCTTGAATAATCTATCG 1543  
DB 300 uThrArgProGlyTyrAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrAr 320  
QY 1544 CCAGGCTCTGGTGGCGGGTCTCCGGGGCGGCTCAAGACGCCCCACCTAATTTGGCTAAGT 1603  
DB 320 gGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVa 340  
QY 1604 AAGAGAAAGTGTGTCAGGGACCGGAATGAACCCCTCTGTTTTCTTTCAGAGGCTCTTGG 1663

```
Db 340 lArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGl 360
Qy 1664 AGCTTCAGGCGGTACACCCCTTTTGATCCACCTCAGAGCCCCAAAAGCCTCAGTGGC 1723
Db 360 uAlaPheArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAl 380
Qy 1724 TTTGGCCCTTATAGGACAGTCAGCCTTGGATATTAGAAAGCTTCAGAGACTGGAGG 1783
Db 380 aLeuAlaPheIleGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGl 400
Qy 1784 GTTACAGGAGCTGAGTTAGCTGATCTAGTCAGGAGGACAGAAAGTATTACAAAG 1843
Db 400 yLeuGlnGluAlaGlnLeuArgAspLeuValArgGluAlaGlnLysValTyrTyrArg 420
Qy 1844 GAGACAGAAAGAAAGGAAACAAAGAAAGAGAGAGAAAGAGGAAAGGAGGAAAG 1903
Db 420 gGluThrGluGluGluLysGluGlnArgLysGluLysGluArgGluGluArg 440
Qy 1904 ACGTAATAACGGAAGAGAAATTTGACTAAGATCTTGGCTGCAAGTGTGTAAGGAA 1963
Db 440 gArgAspArgGlnGluLysLeuThrLysIleLeuAlaValValGluGlyLy 460
Qy 1964 AAGCAATACGAAAGAGAGAGATTTTAGGAAAATTAGTCAGGCCCTAGACAGTCAGG 2023
Db 460 sSerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGl 480
Qy 2024 GAACCTGGGCAATAGAGCCCACTCCGACAAAGGACCAATGTGCATATTGTAAAGAAAGG 2083
Db 480 yAsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGl 500
Qy 2084 ACCTGGGCAAGGAATGCCCCAAGAGGAAACAAAGGACCAAGGATCCTAGCTCTAGA 2143
Db 500 yHisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGl 520
Qy 2144 AGAAGATAAGAT 2156
Db 520 uGluAspLysAsp 524

RESULT 10
US-08-929-967-7
; Sequence 7, Application US/08929967
; Patent No. 5891637
; GENERAL INFORMATION:
; APPLICANT: Rupert, Siegfried J.W.
; TITLE OF INVENTION: Construction of Full-Length cDNA Libraries
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/929,967
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1035R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 amino acids
```

```
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-929-967-7

Alignment Scores:
Pred. No.: 2,2e-205 Length: 665
Score: 2500.50 Matches: 460
Percent Similarity: 83.5% Conservative: 93
Best Local Similarity: 69.5% Mismatches: 106
Query Match: 17.1% Indels: 3
DB: 1 Gaps: 3

US-10-723-552-3 (1-8132) x US-08-929-967-7 (1-665)

Qy 2514 TTGACCCCTCCAATTAGATGACGAATATCGACTATATCTCTCCCTAGTAAAGCCGTGATCAA 2573
Db 1 MetProLeuAsnIleGluAspGluHisArgLeuHisGluThrSerLysGluProAspVal 20
Qy 2574 AATATA--CAATTCTGGTTGGAAACAGTTTCCCAAGCCTGGCCAGAAACCCGAGGGATG 2630
Db 21 SerLeuGlySerThrTrpLeuSerAlaPheProGlnAlaTrpAlaGluThrGlyGlyMet 40
Qy 2631 GGTTCGGCAAGCAAGTTCCTCCCAAGTATTCAACTGNAAGCCAGTCCACACCAAGTG 2690
Db 41 GlyLeuAlaValArgGlnAlaProLeuIleIleProLeuLysAlaThrSerThrProVal 60
Qy 2691 TCAGTCAGACAGTACCTCCCTTGAGTAAAGAGCTCAAGAAAGAAATTCGGCCGCGATGTCAA 2750
Db 61 SerIleLysGlnTyrProMetSerGlnGluAlaArgLeuGlyIleLysProHisIleGln 80
Qy 2751 AGATTAATCAACAGGGCATCTCTAGTCTCTGTCCTCAATCTCCCTGGAAATCTCCCTGCTA 2810
Db 81 ArgLeuLeuAspGlnGlyIleLeuValProCysGlnSerProTrpAsnThrProLeuLeu 100
Qy 2811 CCGTTTAGAAGCCTGGGACTAATGACTATCGACAGTACAGCACTTGAGAGAGTCAAT 2870
Db 101 ProIleLysLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsn 120
Qy 2871 AAACGGGTGAGGATATACACCAACAGTCCCGAACCTTATAACCTCTTGTGTGCTCTC 2930
Db 121 LysArgValGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerGlyLeu 140
Qy 2931 CCACCCCAACGGAGCTGGTATACAGTATTGGACTTAAAGGATGCCTTCTTCTCCCTGAGA 2990
Db 141 ProProSerHisGlnTrpTyrThrValLeuAspLeuLysAspAlaPheCysLeuArg 160
Qy 2991 TTACACCCCACTAGCCAAACCACTTTTGGCTTCGATGAGAGATCCAGTACGGGAAGA 3050
Db 161 LeuHisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGluMetGlyIle 180
Qy 3051 ACGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCGCAGCATCTTT 3110
Db 181 SerGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPhe 200
Qy 3111 GACGAAGCCCTACACAGACACCTGGCCAACTTCAGGATCCAAACCCCTCAGGTGACCCCTC 3170
Db 201 AspGluAlaLeuHisArgAspLeuAlaGlyPheArgIleGlnHisProAspLeuLeuLeu 220
Qy 3171 CTCCTAGCTAGTACCTGCTCTTGGCGGGAGCCCAACAAACAGACACTGCTTAGAAGGC 3230
Db 221 LeuGlnTyrValAspAspLeuLeuAlaIleSerSerSerGluLeuAspCysGlnGlnGly 240
Qy 3231 ACGAAGGCACTACTGCTGAAATTGCTGACTAGGCTCAGAGCCCTCTGTAAGAAGGCC 3290
Db 241 ThrArgAlaLeuLeuGlnThrLeuGlyAspLeuGlyTyrArgAlaSerAlaLysLysAla 260
Qy 3291 CAGATTTGCAGGAGAGAGTAAACATCTTGGGGTACAGTTTGGGACCGGAGCGAGTGG 3350
Db 261 GlnIleCysGlnLysGlnValLysTyrLeuGlyTyrLeuLeuLysGluGlyGlnArgTrp 280
Qy 3351 CTGACGGAGGACCGAAGAAACTGTAGTCCAGATACCGGCCCCCAACCAACAGCAACAA 3410
Db 281 LeuThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGln 300
```

```
QY 3411 ATGAGAGCTTTTGGGACAGCTGGATTTTGCAGCTGTGGATCCCGGGTTCGGACC 3470
Db 301 LeuArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpLeuPheAlaGlu 320
QY 3471 TTAGCAGCCCACTCTACCCGCTAAACCAAGAAAGGGAAATTCCTCGGCTCTCGAG 3530
Db 321 MetAlaAlaProLeuTyPProLeuThrLysThrGlyThrLeuPheAsnTrpGlyProAsp 340
QY 3531 CACAGAGGCAATTTGATGTATCAAAAGCCCTGTGTAGCGCACTGTCTGTGGCCCTC 3590
Db 341 GlnGlnLysAlaTyGlnGluLeuLysGlnAlaLeuThrAlaProAlaLeuGlyLeu 360
QY 3591 CTTGAGCTAACTAAACCTTTATCCCTTATGTGGATGAGCGTAAGGAGTAGCCGGGA 3650
Db 361 ProAspLeuThrLysProPheGluLeuPheValAspGlnLysGlnGlyTyAlaLysGly 380
QY 3651 GTTTTAAACCAACCCCTAGACCATGAGAGACCTGTCCCTACCTGTCAAAAGAGCTC 3710
Db 381 ValLeuThrGlnLysLeuGlyProTrpArgProValAlaTyLeuSerLysLeu 400
QY 3711 GATCTGTAGCCAGTGTGGCCCATATGCTGAAGCTATCGAGCTGTGGCCATCTG 3770
Db 401 AspProValAlaAlaGlyTrpProCysLeuArgMetValAlaAlaIleAlaValLeu 420
QY 3771 GTCAAGGACGCTGACAAATTCATTTGGGACAGATATACTGTATAGCCCCCATGCA 3830
Db 421 ThrLysAspAlaGlyLysLeuThrMetGlyGlnProLeuValIleLeuAlaProHisAla 440
QY 3831 TTGGAGAACTCGTTCCGAGCCCCAGACCGATGATGACCAACGCCGCCCATGACCCAC 3890
Db 441 ValGluThrLeuValLysGlnProProAspArgTrpLeuSerAsnAlaArgMetThrHis 460
QY 3891 TATCAAGCCTGCTTCTC---ACAGAGAGGGTCACGTTGCTCCACAGCGCTCTCAAC 3947
Db 461 TyrGlnAlaLeuLeuLeuAspThrAspArgValGlnPheGlyProValValAlaLeuAsn 480
QY 3948 CTTGCCACTCTTCTGCTGAAGAGACTGTGAACAGTGAATCATGATTCGCATCACTA 4007
Db 481 ProAlaThrLeuLeuPro---LeuProLysGluGlyLeuGlnHisAspCysLeuAspIle 499
QY 4008 TTGATTGAGGAGACTGGGGTCCGCAAGACCTTACAGACATACCGCTGACTGGAGAAGTG 4067
Db 500 LeuAlaGluAlaHisGlyThrArgSerAspLeuThrAspGlnProLeuProAspAlaAsp 519
QY 4068 CTAACCTGTTCTACTGAGGAAGCAGCTATGTGTGGAAGGTAAAGGATGGTGGGGCG 4127
Db 520 HisThrTrpTyThrAspGlySerSerPheLeuGlnGluGlyGlnArgLysAlaGlyAla 539
QY 4128 GCGTGTGTGACGGGACCCGACGATCTGCCAGCAGCTGCGGAGGAGAACTTCAGCA 4187
Db 540 AlaValThrThrGluThrGluValIleTrpAlaArgAlaLeuProAlaGlyThrSerAla 559
QY 4188 CAAAAGGCTGAGTCACTGCGCCCTCACCAAGCTTTGCGGCTGCGCAAGGAAATCCATA 4247
Db 560 GlnArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLysLeu 579
QY 4248 AACATTTATAGGACAGCAGGTATGCTTTGCGACTGCACACGTACATGGGGCCATCTAT 4307
Db 580 AsnValTyThrAspSerArgTyAlaPheAlaThrAlaHisIleHisGlyGluIleTy 599
QY 4308 AAACAAGGGGTGTCTTACTCAGCAGGAGGGAATTAAGAACAAAGGAATTTCTA 4367
Db 600 ArgArgArgGlyLeuLeuThrSerGluGlyGluIleLysAsnLysGlyGluIleLeu 619
QY 4368 AGCTATTAGAAGCCGTACATTTACCAAAAGGCTAGCTATTATACACTGCTCTGGACAT 4427
Db 620 AlaLeuLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysProGlyHis 639
QY 4428 CAGAAAGCTAAGATCTCATATCCAGAGGAAACCGAGATGCTGACCGGGTTGCCAGACG 4487
Db 640 GlnLysGlyAsnSerAlaGluAlaArgGlyAsnArgMetAlaAspGlnAlaAlaArgLys 659
```

```
QY 4488 GCAGCC 4493
Db 660 AlaAla 661
RESULT 11
US-09-111-085-4
; Sequence 4, Application US/09111085
; Patent No. 610034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; FILE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Porcine retrovirus
US-09-111-085-4
Alignment Scores:
Pred. No.: 1,2e-198 Length: 657
Score: 2422.00 Matches: 457
Percent Similarity: 80.2% Conservative: 68
Best Local Similarity: 69.8% Mismatches: 96
Query Match: 16.5% Indels: 34
DB: 2 Gaps: 9
US-10-723-552-3 (1-8132) x US-09-111-085-4 (1-657)
QY 5620 ATGCATCCCACTTAACCGCGCCACCTCCCGATTCGGGTGCAAGCGAAAGACTG 5679
Db 1 MetHisProThrLeuSerTrpArgHisLeuProThrArgGlyGluProLysArgLeu 20
QY 5680 AAAATCCCCTTAAGCTTCGCTCCATCGCGTGGTCTTACTCTGTCATAAATCTCTCAG 5739
Db 21 ArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrProGln 40
QY 5740 ACTAATGTGATGCGCATAGAGACAGCTGAACCTCCATCAACACACTCAAGGGAGGCTCTTTA 5859
Db 41 AlaSerSerLysArgLeuLeuAspSerSerAsnProHisArgProLeuSerLeuThrTrp 60
QY 5800 TTAATTAATGACTCCGCGCAGGTATTATATCAACAACTCAAGGGAGGCTCTTTA 5859
Db 61 LeuIleIleAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProArg 80
QY 5860 GGAACCTGTGGCTGATCTATACGTTTGCTCAGATCAGTATTCTAGTCTGACCTCA 5919
Db 81 GlyThrTrpTrpProGluLeuHisPheCysLeuArgLeuLeuAsnProAlaValLysSer 100
QY 5920 -----CCCCAGATATCTCCATGCTCAGGATTTTATGTTTGCACGAGCACCAATAAT 5976
Db 101 ThrProAsnLeuValArgSerTyGlyPheTyCysCysProGlyThrGluLysGlu 120
QY 5977 GGAACAATTCGCGAATCCAGAGATTTCTTTGTAAACAATGGAACCTGTGAACCTCT 6036
Db 121 ----LysTyrcysGlyGlySerGlyGluSerPheCysArgArgTrpSerCysValThrSer 139
QY 6037 AATGATGATATTGGAATGGCAACCTCTCAGCAGATAGGTAAGTTTCTTATGTC 6096
Db 140 AsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
QY 6097 AACACCTATACAGCTCTGGCAATTTAATTAATCACTGACCTGGATAGAACTGGNAGCCCC 6156
Db 160 Asn-----SerGlyProGlyLysTyLysValMetLysLeuTyLysAspLys 175
QY 6157 AAGTGTCTCTTTCAGACCTAGATTACCTAAATAAAGTTTCTACTGAGAAAGGAAACAA 6216
```



Db 61 LeuIleleAspProaspThrGlyValThrValAsnSerThrArgGlyValAlaProArg 80  
QY 5860 GGAACCTGGTGGCTGATCATACGTTTGCTCAGATCAGTATATCTAGTCTGACCTCA 5919  
Db 81 GlyThrTrpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100  
QY 5920 ---CCCCAGATATCCATGCTCAGGATTTTATGTTTGGCCAGGACCAACAATAAT 5976  
Db 101 ThrProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGlu 120  
QY 5977 GGAACCATTCGCGAATCCAGAGATTCTTTTGTAAACAATGGAACCTGTGAACCTCT 6036  
Db 121 ---LysTyrCysGlySerGlyGluSerPheCysArgArgTrpSerCysValThrSer 139  
QY 6037 AATGATGATATTTGGAATAGCCCAACCTCTCAGCAGATAGGTAAGTTTCTTATGTC 6096  
Db 140 AsnAspGlyAspTrpLysTrpProLysLeuGluAspArgValLysPheSerPheVal 159  
QY 6097 AACACCTATACCACTCTGACAAATTAATTACTGACCTGGATTAGAACTGAAGCCCC 6156  
Db 160 Asn-----SerGlyProGlyLysValMetLysLeuTyrLysAspLys 175  
QY 6157 AAGTGCTCTCTTCAGACCTAGATTACTTAAATAAGTTTCTCAGAAAGGAAACAA 6216  
Db 176 SerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGln 195  
QY 6217 GAAATATCTTAAATGGGTAATGATGATCTCTGGGGAATGATATAT-----TATGGA 6270  
Db 196 GluAsnIleGlnLysTrpLysAsnGlyMetSerTrpGlyIleValPheTyrLysTyrGly 215  
QY 6271 GGCTCGGTGTAACACACAGCTCCATCTTAACCTTTCGCTCAAAATAAACACAG---CTG 6327  
Db 216 GlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 232  
QY 6328 GAGCCTCCAATGCTATAGACCAATAACGGTCTTGACGGTCAAAGACCCCAACCCAA 6387  
Db 233 GluProProValAlaValGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 252  
QY 6388 GGAACCA-----GGA 6396  
Db 253 GluProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrGln 272  
QY 6397 CCATCCTCTAACATACTCTGATCAGACCCCACTCAGTCT-----AACACG----- 6444  
Db 273 ProProSerAsnGlyThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 292  
QY 6445 ---ACGACTAAATGGGGCAAACTTTTATGCTCATCCAGGAGCTTTTCAAGCTCTT 6501  
Db 293 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 312  
QY 6502 AACTCCACGACTCCAGAGGCTACTCTTCTTGTGGCTATGCTTAGCTTCGGGCCCACT 6561  
Db 313 AsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProPro 332  
QY 6562 TACTATGAAGATGGCTAGAGGGAATTCATGTGACAAAGACATAGACCAAA 6621  
Db 333 TyrTyrGluGlyMetAlaLysGluGlyLysPheAsnValThrLysGluHisArgAsnGln 352  
QY 6622 TGACATGGGGATCCCAAAATAAGCTTACCCTTACTGAGGTTTCTGGAAAAGGCACCTGC 6681  
Db 353 CysThrTrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCys 372  
QY 6682 ATAGGAAGGTTCCCAATCCCAACCAACACTTTGTAAACACACTGAAGCCTTTAATCAA 6741  
Db 373 IleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 392  
QY 6742 ACCTCTGAGAGTCAATATCGTACCTGGTTATGACAGGTGGTGGCATGTAATCTGGA 6801  
Db 393 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpAlaCysAsnThrGly 412  
QY 6802 TTAACCCCTTGTTGTTCCACTTGGTTTAAACCAACTAAAGATTTTTCATTATGGTC 6861

Db 413 LeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheCysValMetVal 432  
QY 6862 CAAATTTGTTCCCGAGTGTATTACTATCCGAAAAAGCAATCTTGTATGATATGACTAC 6921  
Db 433 GlnIleValProArgValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyr 452  
QY 6922 AGAAATCATCGCAAAAGAGAGAACCATATCTCTGACACTTCTGATGCTCGGACTT 6981  
Db 453 ArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeu 472  
QY 6982 GGAAGTGGCAGCAGCTGTAGGAAACAGAACAGCTGCCCTGTGTACGGGACCAACAGCTA 7041  
Db 473 GlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 492  
QY 7042 GAAACAGGACTTAGTAACCTACATCGAATTGTAAAGAGATCTCAAGCCCTAGAAAAA 7101  
Db 493 GluLysGlyLeuGlyLeuHisAlaIleMetThrGluAspLeuArgAlaLeuGluGlu 512  
QY 7102 TCTGTCAAGTAACTCGAGGAATCCCTAACTCTTATCTGAAGTAGTCTTACAGAATAGA 7161  
Db 513 SerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArg 532  
QY 7162 AGAGGTTTAGATTTATTTCTTAAAGAGGAGGATTATGTGTAGCTTGAAGAGGAA 7221  
Db 533 ArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGlu 552  
QY 7222 TGCTGTTTTTATGTGATCATTCAGGGGCCATCAGAGACTCCATGATGAACAGCTTAGAGAA 7281  
Db 553 CysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGlu 572  
QY 7282 AGTTTCGAGAAGCGTCGAAGGAAAAAGAACTACTCAAGGGTGGTTTGAGGGATGGTTC 7341  
Db 573 ArgLeuGluArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPhe 592  
QY 7342 AACAGTCTCTTTGGTGGCTACCTACTTTTCTGTCTTAAACAGGACCTTTAATAGTCTC 7401  
Db 593 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValLeu 612  
QY 7402 CTCCTGTTACTCAGCTTGGCCCATGATATTAACAAGTTAATTCCTTCAATAGAGAA 7461  
Db 613 LeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGlu 632  
QY 7462 CGAATAAGTGCAGTCCAGATCATGTGTACTTTAGACAACAGTACCACAA 7506  
Db 633 ArgValSerAlaValGlnIleMetValLeuArgGlnTyrGln 647

## RESULT 13

US-09-171-553B-10  
Sequence 10, Application US/09171553B  
Patent No. 6756227  
GENERAL INFORMATION:  
APPLICANT: GALBRAITH, DANIEL N.  
APPLICANT: HAWORTH, CHRISTINE  
APPLICANT: LILES, GILLIAN M.  
APPLICANT: SMITH, KENNETH T.  
TITLE OF INVENTION: PORCINE RETROVIRUS  
FILE REFERENCE: CFV-5.01  
CURRENT APPLICATION NUMBER: US/09/171,553B  
CURRENT FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: PCT/GB97/01087  
PRIOR FILING DATE: 1997-04-18  
PRIOR APPLICATION NUMBER: GB 9702668.6  
PRIOR FILING DATE: 1997-02-10  
PRIOR APPLICATION NUMBER: GB 9608164.1  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 10  
LENGTH: 657  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Amino acid sequence



OTHER INFORMATION: of ENV region "Raji"  
US-09-171-553B-10

Alignment Scores: Pred. No.: 2 5e-196 Length: 657  
Score: 2395.00 Matches: 451  
Percent Similarity: 80.3% Conservative: 75  
Best Local Similarity: 68.9% Mismatches: 95  
Query Match: 16.4% Indels: 34  
DB: 2 Gaps: 10

US-10-723-552-3 (1-8132) x US-09-171-553B-10 (1-657)

```

QY 5620 ATGCATCCCAAGTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAGCGAAAGACTG 5679
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1 MetHisProThrLeuSerTrpAlaGluProThrArgGlyGlyGluProLysArgLeu 20
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5680 AAATCCCTTAAGCTCGCTCCATCGCTGCTTCTTCTCTCTCTCTCTCTCTCTCTCAG 5739
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 21 ArgileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrIleThrProGln 40
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5740 ACTAATGGTATGGCATAGAGACAGAGCTGAACTCCCATAAACCTTATCTCTCAGCTGG 5799
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 41 AlaSerSerLysArgLeuIleAspSerSerAsnProHisArgProLeuSerProThrTrp 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5800 TTAATTACTGATCCCGGACAGGTATTAAATCAACAACACTCAAGGGGAGGCTCCTTTA 5859
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 LeuIleAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProArg 80
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5860 GGAACCTGGTGGCTGATCTATAGCTTGCTCAGATCAGTATTCTCTAGCTGACCTCA 5919
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 81 GlyThrTrpTrpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5920 ---CCCCAGATATCCTCCATGTCTCAGGATTTTATGTTTGGCCAGGACCAACAATAAT 5976
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 101 ThrProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGlu 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5977 GGAACAATTCGCGAATCCAGAGATTTCTTTTGTAAACAATGGAACCTGTGTAACCTCT 6036
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 ---LysTyrCysGlyGlySerGlyGluSerPheCysArgArgTrpSerCysValThrSer 139
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6037 AATGATGGATATGGAATGGCCAACTCTCAGCAGATAGGTAAGTTTCTTATGTC 6096
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 140 AsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6097 AACACCTATACAGCTCTGACAAATTTAATACCTGACCTGGATTGAAGTGAAGCCCC 6156
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 160 Asn-----SerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAspLysSer--- 176
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6157 AAGTGTCTCTCTCAGACCTAGATTACCTAAAAATAAGTTTCACTGAGAAAGGAAACAA 6216
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 177 ---CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGln 195
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6217 GAAATATCTCTAAATGGTAAATGTTATGCTCTGGGGAATGTTATAT-----TATGGA 6270
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 196 GluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTyrLysTyrGly 215
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6271 GGCTCGGTAACAACACGAGCTCCATTCTAATCTTTCGCTCAAAATAAACCCAG-----CTG 6327
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 216 GlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 232
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6328 GAGCTCCCAATGGCTTAGACCAATAAGCTTTCAGCGGTCAAGAGACCCCAACCCAA 6387
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 233 GluProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 252
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6388 GGACCA-----GGA 6396
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 253 GluProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrGln 272
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6397 CCATCTCTTAACATACTTCTGATCAGACCCCACTGAGTCT-----NACAGC----- 6444
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 273 ProProSerAsnSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 292
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 14

```

QY 6445 ---ACGACTAAATGGGGCAAAACATTTTAGCTCATCCAGGAGGCTTTTCAAGCTCTT 6501
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 293 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 312
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6502 AACTCCAGACTCCAGAGGCTACTCTTCTTGTGGCTATGCTTAGCTTCGGGCCCACT 6561
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 313 AsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProPro 332
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6562 TACTATGAGGAATGCTAGAACGAGGGAATTCATATGTGTGACAAAGAACATAGAGACAA 6621
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 333 TyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGln 352
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6622 TGACATGGGGATCCCAATAAAGCTTACCTTACTGAGGTTCCTGGAAGGACCACTGC 6681
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 353 CysThrTrpGlySerArgAsnLysLeuThrLeuGluValSerGlyLysGlyThrCys 372
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6682 ATAGGAAGGTTCCCATCCACCAACACCTTTGTAAACCACTAGAGCTTTAATCAA 6741
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 373 IleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 392
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6742 ACCTCTGAGAGTCAATATCTGTTACCTGGTTATGACAGGTGGTGGGCATGTAATCTGGA 6801
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 393 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGly 412
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6802 TTAACCCCTTGTGTTTCCACCTTGGTTTAAACCAACTAAAGATTTTTCATTTATGGTC 6861
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 413 LeuThrProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysValMetVal 432
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6862 CAATTTGTTCCCGAGTGTATTACTATCCGGAAGAAAGCAATCTCTGATGAATATGACTAC 6921
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 433 GlnIleValProArgValTyrTyrHisProGluValValLeuAspGluTyrAspTyr 452
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6922 AGAATATCATCGACAAAGAGAGAACCATATCTCTGACACTTCTGTGTGCTCGGACTT 6981
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 453 ArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeu 472
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 6982 GGAGTGGCAGCAGGTGTAGGAACAGGAACAGCTGCTGTGTCTACCGGACCAACAGCAGCTA 7041
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 473 GlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 492
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7042 GAAACAGGACTTGTATTAACCTACATCGAATTTGTAACAGAGATCTCCAGCCCTTAGAAAAA 7101
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 493 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlu 512
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7102 TCTGTGAGTAACCTGAGGAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGATAGA 7161
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 513 SerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArg 532
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7162 AGAGGTTAGATTATTATTCTTAAAGAGGAGGATTATGTAGCTTTGAAGGAGGAA 7221
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 533 ArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGlu 552
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7222 TGCTGTTTTTATGTGATCATTCAGGGGCGCATCAGAGACTCCATGAACAGCTTAGAGAA 7281
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 553 CysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLys 572
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7282 AGGTTGAGAGCGTCTGAGGGAAGGAAACTACTCAAGGGTGTGTTGAGGAGTGTCTTC 7341
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 573 LysLeuGluArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPhe 592
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7342 AACAGTCTCTTTTGGTGGTACCTTACTTCTTCTTAAACAGGACCTTTAATAGTCCCTC 7401
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 593 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu 612
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7402 CTCCTGTTACTCAGAGTGGGCCATGTATTATTAAAGTTTAATTCCTTCTTATTAGAGAA 7461
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 613 LeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGlu 632
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7462 CGAATAAGTGCAGTCCAGATCATGTACTTACACACAGTACCAA 7506
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 633 ArgValSerAlaValGlnIleMetValLeuArgGlnTrpGln 647
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
US-09-376-781-3
; Sequence 3, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papi T.
; APPLICANT: Patience, Clive
; APPLICANT: Anderson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Amino acid
; OTHER INFORMATION: sequence of a PERV env polypeptide previously
; OTHER INFORMATION: described
US-09-376-781-3

Alignment Scores:
Pred. No.: 1,01e-192 Length: 656
Score: 2353.00 Matches: 450
Percent Similarity: 79.8% Conservative: 74
Best Local Similarity: 68.5% Mismatches: 96
Query Match: 16.1% Indels: 37
DB: 2 Gaps: 10

US-10-723-552-3 (1-8132) x US-09-376-781-3 (1-656)
QY 5620 ATCCATCCCAAGTTAAACCGCGCCACCTCCGATTCGGGTGGAAAGCCGAAAGACTG 5679
Db 1 MetHisProThrLeuSerArgGHisLeuProThrArgGlyGlyGluProLysArgLeu 20
QY 5680 AAAATCCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCAATAACCTCTCAG 5739
Db 21 ArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrIleThrProGln 40
QY 5740 ACTAATGGTATGCGCATAGGAGACGCTGAACTCCCATAAACCTTTATCTCTCACCCTGG 5799
Db 41 AlaSerSerLysArgLeuIleAspSerSerAsnProHisArgProLeuSerLeuThrTrp 60
QY 5800 TTAATTAAGTACCTCCGCGCAGGTATTAATATCAACAACACTCAAGGGAGGCTCCTTTA 5859
Db 61 LeuIleLeaspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProArg 80
QY 5860 GGAACCTGGTGGCTGATATATACGTTTGCTCAGATTCAGTATTCTAGTCTGACCTCA 5919
Db 81 GlyThrTrpTrpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
QY 5920 ---CCCCAGATATCTCCAGTCTCAGGATTTATGTTTGGCCAGGACCCACCAATAAT 5976
Db 101 ThrProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGlu 120
QY 5977 GGAARAACATTCGGAATCCAGAGATTTCTTTGTAAACAATGGAAGTCTGTAACTCT 6036
Db 121 ---LysTyrCysGlyGlySerGlyGluSerPheCysArgArgTrpSerCysValThrSer 139
QY 6037 AATCATGGATTAATGGAATCCCAACCTCTCAGCAGGATAGGATAGGTTTCTTATGTC 6096
Db 140 AsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
QY 6097 AACACCTATACAGCTCTGACAAATTAATTAATCTACCTGGATTAAGAACTGGAAGCCCC 6156
Db 160 Asn-----SerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAspLysSer--- 176

6157 AAGTGCTCTCCTTCAGACACCTAGATTACCTAAATAAAGTTTCTACCTAGAAAGGAAACAA 6216
177 ---CysSerProSerAspLeuaspTyrLeuLysIleSerPheThr--GluArgLysThrG 195
6217 GAAATATCTCTAAATGGGTAAATGTATGCTCTTGGGGGAATGTATAT-----TATGGA 6270
195 LysLysTyrSerLysValaspLysTrpTyrGluLeuGlyAsnSer-PheLeuLeuTyrGly 214
6271 GGCTCGGTAAACAACACGAGCTCCATCTTAACCTATTTCGCTCAATAAATACACAG---CTG 6327
215 GlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 231
6328 GAGCTCTCAATGGCTATAGGACCAAAATACGGTCTTGACGGGTCAAAAGACCCCAACCAAA 6387
232 GluProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 251
6388 GGACCA-----GGA 6396
252 GluProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrGln 271
6397 CGATCCTCTAACAATACTTCTGGATCAGACCCCACTGAGTCT-----AACAGC----- 6444
272 ProProSerAsnSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 291
6445 ---ACGACTAAATGGGGCAAAACTTTTGTAGCTCATCCAGGGAGCTTTTCAAGCTCTT 6501
292 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 311
6502 AACTCCAGACTCCAGAGGCTACCTCTTCTTGTGTGCTATGCTTGTAGCTCGGGCCACCT 6561
312 AsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProPro 331
6562 TACTATGAGGATGCTAGAGGGAAATTCAAATGTGACAAAGAAACATAGAGACCAA 6621
332 TyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGln 351
6622 TGACATGGGGATCCCAAAATAAGCTTACCTTACTGAGGTTTCTGGAAGGACACCTGC 6681
352 CysThrTrpGlySerArgAsnLysLeuThrLeuGluValSerGlyLysGlyThrCys 371
6682 ATAGGAAAGGTTCCCATCCCAACCAACCTTTGTAAACCACTGAAGCCTTTTAAATCAA 6741
372 IleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 391
6742 ACCTCTGAGAGTCAATATCTGCTACCTGTTATGACAGGTGGTGGCATCTATATCTGGA 6801
392 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGly 411
6802 TTAACCCCTTGTGTTCCACCTTGGTTTAAACCAAACTAAAGATTTTTCATTATGGTC 6861
412 LeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheCysValMetVal 431
6862 CAAATTTGTTCCCGAGTGATTAATCTATCCGAAAAAGCAATCTCTGATGAATATGACTAC 6921
432 GlnIleValProArgValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyr 451
6922 AGAAATCATCGCAAAAGAGAGAACCCATATCTCTGACCTTCTGCTGTGCTCGGACTT 6981
452 ArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeu 471
6982 GGAGTGGCAGCAGGTGTAGGAAACAGAAACAGCTGCGCTGGTACGGGACCCACAGCAGCTA 7041
472 GlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 491
7042 GAAACAGGACTTATAGTAACTTACATCGAATGTTTAACAGAGATCTCCAAGCCCTAGAAAAA 7101
492 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlu 511
7102 TCTGTCAAGTAACTCGAGGAACTCCCTAACTCTTATCTGAAGTGTCTTACAGAGATAGA 7161
512 SerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArg 531
7162 AGAGGCTTAGATTATTATTCTTAAAGAGGAGGATTTATGTGTAGCTTTGAAAGGAGAA 7221
```

Db 532 ArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGlu 551  
QY 7222 TGCTGTTTATGTGATCATTCAGGGCCCATCAGAGACTCCATGAAACAGCTTAGAGAA 7281  
Db 552 CysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAnLysLeuArgLys 571  
QY 7282 AGTTGAGAGCGTCCGAGGGAAGAAAGAACTACTCAAGGGTGGTTGAGGATGCTTC 7341  
Db 572 LysLeuGluArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPhe 591  
QY 7342 AACAGTCTCTTGGTGGTACCTACTTCTCTTAAACAGGCTTAAAGTCTTC 7401  
Db 592 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu 611  
QY 7402 CTCTGTACTACAGTTCGGCCCATGATTATTAAACAAGTAATTCGCTTCATTAGAGAA 7461  
Db 612 LeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGlu 631  
QY 7462 CGAATAAGTCAGTCCAGATCATGGTACTTAGACAACAGTACCAG 7506  
Db 632 ArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGln 646

## RESULT 15

US-09-171-553B-6  
; Sequence 6, Application US/09171553B  
; Patent No. 6756227  
; GENERAL INFORMATION:  
; APPLICANT: GALBRAITH, DANIEL N.  
; APPLICANT: HAWORTH, CHRISTINE  
; APPLICANT: LEE, GILLIAN M.  
; APPLICANT: SMITH, KENNETH T.  
; TITLE OF INVENTION: PORCINE RETROVIRUS  
; FILE REFERENCE: CFV-5.01  
; CURRENT APPLICATION NUMBER: US/09/171,553B  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: PCT/GB97/01087  
; PRIOR FILING DATE: 1997-04-18  
; PRIOR APPLICATION NUMBER: GB 9702668.6  
; PRIOR FILING DATE: 1997-02-10  
; PRIOR APPLICATION NUMBER: GB 9608164.1  
; PRIOR FILING DATE: 1996-04-19  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 656  
; TYPE: PRT  
; ORGANISM: Porcine retrovirus  
US-09-171-553B-6

Alignment Scores:  
Pred. No.: 1,01e-192 Length: 656  
Score: 2353.00 Matches: 450  
Percent Similarity: 79.8% Conservative: 74  
Best Local Similarity: 68.5% Mismatches: 96  
Query Match: 16.1% Indels: 37  
DB: 2 Gaps: 10

US-10-723-552-3 (1-8132) x US-09-171-553B-6 (1-656)

QY 5620 ATGCATCCACAGTTAAACCGGCCCATCTCCGATTCGGGGTGAAGCCGAAAGACTG 5679  
Db 1 MetHisProThrLeuSerArgHisLeuProThrArgGlyGlyLeuProLysArgLeu 20  
QY 5680 AAAATCCCTTAAGCTTCGCTCCATCGGTGGTTCCTTACTCTGTCATTAACCTCTAG 5739  
Db 21 ArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrIleThrProGln 40  
QY 5740 ACTAATGGTATGCGCATAGGACAGCTGAACTCCCAATAAACCTTATCTCTACCTGG 5799  
Db 41 AlaSerSerLysArgLeuIleAspSerSerAsnProHisArgProLeuSerLeuThrTrp 60  
QY 5800 TTAATTAAGTCCGGCACAGGTATTAAATATCAACACACTCAAGGGGAGGCTCCTTTA 5859

Db 61 LeuIleIleAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProArg 80  
QY 5860 GGAACCTGTGGCTCATCTATAGTTTGCCTCAGATCAGTTATTCTTCTGCTGACCTCA 5919  
Db 81 GlyThrTrpProGlnLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100  
QY 5920 ---CCCCAGATATCTCCATGCTCAGGATTTTATGTTTGCCTCAGGACCCACCAATAA 5976  
Db 101 ThrProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGlu 120  
QY 5977 GGAACAATTCGGGAATCCAGAGATTTCTTTTAAACAATGGAACCTGTGAACCTCT 6036  
Db 121 ---LysTyrCysGlyGlySerGlyGluSerPheCysArgArgTrpSerCysValThrSer 139  
QY 6037 AATGATGGATATTGGAATGCGCAACCTCTCAGCAGGATAGGTAAGTTTCTTCTTATGTC 6096  
Db 140 AsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159  
QY 6097 AACACCTATACAGCTCTCGACAATTTAATTACTGACCTGGATTAGAACCTGGAAGCCCC 6156  
Db 160 Asn-----SerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAspLysSer--- 176  
QY 6157 AAGTGTCTCTTCAGACCTAGATTACTTAAATAAGTTTCTCAGAGAAAGAAACAA 6216  
Db 177 ---CysSerProSerAspLeuAspTyrLeuLysIleSerPheThr--GluArgLysThrG 195  
QY 6217 GAAATATCTTAAATGGGTAAATGGTATGCTCTGGGGAATGGTATAT-----TATGGA 6270  
Db 195 LysLysTyrSerLysValAspLysTrpTyrGluLeuGlyAsnSer--PheLeuLeuTyrGly 214  
QY 6271 GGCTCGGGTAAACCAACAGGCTCCATTCTAACTATTTCGCTCAAAATAAACACAG---CTG 6327  
Db 215 GlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 231  
QY 6328 GAGCTTCAATGGCTATAGGACCAATACGGTCTTGACGGGTCAAGAGACCCCAACCCAA 6387  
Db 232 GluProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyProAlaLeu 251  
QY 6388 GGACCA-----GGA 6396  
Db 252 GluProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrGln 271  
QY 6397 CCATCTCTTAACATAACTTCTGGATCAGACCCCATCTGAGTCT-----AACAGC----- 6444  
Db 272 ProProSerAsnSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 291  
QY 6445 ---ACGACTAAATGGGGCAAACTTTTACCTCATCCAGGAGCTTTTCAAGCTCTT 6501  
Db 292 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 311  
QY 6502 AACTCCAGACTCCAGAGGCTACCTTCTTCTTGGCTATGCTTAGCTTCGGGCCCCACCT 6561  
Db 312 AsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProPro 331  
QY 6562 TACTATGAAGGAATGCTAGAAGAGGAAATCAATGTGACAAAGAAACATAGAGACCAA 6621  
Db 332 TyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGln 351  
QY 6622 TCACATGGGATCCCAAAATAAGCTTACCTTACTGAGTTTCTGGAAGGACCTGC 6681  
Db 352 CysThrTrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCys 371  
QY 6682 ATAGGAAGGTTCCCTCCATCCCAACACCTTCTTAACACACACTGAACCTTTAATCAA 6741  
Db 372 IleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 391  
QY 6742 ACCTCTGAGAGTCAATATCTGGTACTGTTATGACAGTGGTGGCATGTAATCTGGA 6801  
Db 392 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGly 411  
QY 6802 TTAACCCCTTGTTTCCACCTTGGTTTAAACCAAACTTAAAGATTTTTCATTATGGTC 6861

Db	412	LeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheCysValMetVal	431
Qy	6862	CAAAATTGTTCCCGAGGTATTACTATCCGAAAGAAACAATCCTTGTGATGAATATGACTAC	6921
Db	432	GlnIleValProArgValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyr	451
Qy	6922	AGAAATCATCGACAAAGAGAGAACCATATCTCTGACACTTGCTGTGATGCTCGGACTT	6981
Db	452	ArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeu	471
Qy	6982	GGAGTGCAGCAGGTGTAGGAAACAGGAACAGCTCCCTCGTTCAGCGGACCACACAGCAGCTA	7041
Db	472	GlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu	491
Qy	7042	GAACAGGACTTAGTAACCTACATCATGTAATGTGTAAACAGAGATCTCCAAGCCCTAGAAAAA	7101
Db	492	GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlu	511
Qy	7102	TCTGTGAGTAACCTGGAGGAATCCCTAACCTCCTTATCTCAGGAGTGCCTACAGAAATAGA	7161
Db	512	SerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArg	531
Qy	7162	AGGGGTAGATTTATTATTCTAAAAAGAGGAGGATTATGTGTAGCCTTTGAAGGAGGAA	7221
Db	532	ArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGlu	551
Qy	7222	TGCTGTTTTTATGTGGATCATTCAGGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAA	7281
Db	552	CysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLys	571
Qy	7282	AGGTTGGAGAGCGCTCGAAGGGAAAGAAACTACTCAAGGGTGGTTGAGGGATGGTTC	7341
Db	572	LysLeuGluArgArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrPhe	591
Qy	7342	AACAGGTCTCTTTGGTTGGCTACCCCTACTTTCTGCTTTAACAGAGCCCTTAATAGTCTC	7401
Db	592	AsnArgSerProTyrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu	611
Qy	7402	CTCCTGTTTACTCACAGTTGGGCCATGTATTATTAAACAAGTTAAATGTGCTTCATTAGAGAA	7461
Db	612	LeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGlu	631
Qy	7462	CGAATAAGTCAGTCCAGATCATCGTACTTAGACAAACAGTACCAA	7506
Db	632	ArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGln	646

Search completed: February 14, 2006, 16:15:06  
Job time : 425.147 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 16:07:51 ; Search time 204.475 Seconds  
(without alignments)  
3323.427 Million cell updates/sec

Title: US-10-723-552-3

Perfect score: 14636

Sequence: 1 GCGTGTGTACGACTGTGG.....CTGTTTGCATCAAAAAAAA 8132

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DRV=xlp  
-Q=/abes/ABSSWEB\_spool/US10723552/runat\_14022006\_125148\_13205/app\_query.fasta\_1  
-DB=Published\_Applications\_AA\_Main -QFMT=fastan -SUFFIX=rapbm -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs02p  
-USER=US10723552 @CGN 1.1.602 @runat\_14022006\_125148\_13205 -NCPU=6 -ICPU=3  
-NO\_WRAP -NEG\_SCORES=0 -WAIT\_DSBLCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA\_Main:

1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3980.5	27.2	1199	4	US-10-677-558-2
2	3425	23.4	638	3	US-09-851-859A-6
3	3412	23.3	638	4	US-10-441-949-32
4	3412	23.3	638	4	US-10-441-949-34
5	3412	23.3	638	4	US-10-441-949-40
6	3412	23.3	638	4	US-10-441-949-42
7	3412	23.3	638	4	US-10-441-949-45
8	3318.5	22.7	678	4	US-10-441-949-36
9	3318.5	22.7	678	4	US-10-441-949-38
10	3147.5	21.5	653	4	US-10-029-656-2
11	3135.5	21.4	653	4	US-10-029-656-4

12	2879.5	19.7	661	4	US-10-441-949-43	Sequence 43, Appl
13	2874.5	19.6	660	4	US-10-441-949-16	Sequence 16, Appl
14	2871.5	19.6	660	4	US-10-441-949-8	Sequence 8, Appl
15	2866.5	19.6	660	3	US-09-851-859A-5	Sequence 5, Appl
16	2851	19.5	678	4	US-10-441-949-12	Sequence 12, Appl
17	2849.5	19.5	660	4	US-10-441-949-10	Sequence 10, Appl
18	2849.5	19.5	660	4	US-10-441-949-18	Sequence 18, Appl
19	2835	19.4	678	4	US-10-441-949-14	Sequence 14, Appl
20	2491	17.0	716	5	US-10-827-498-2	Sequence 2, Appl
21	2487.5	17.0	716	4	US-09-845-157-2	Sequence 2, Appl
22	2487.5	17.0	716	4	US-10-661-819-2	Sequence 2, Appl
23	2485	16.8	678	4	US-10-441-949-26	Sequence 26, Appl
24	2455	16.8	678	4	US-10-441-949-24	Sequence 24, Appl
25	2432	16.6	657	4	US-10-441-949-22	Sequence 22, Appl
26	2432	16.6	657	4	US-10-441-949-30	Sequence 30, Appl
27	2432	16.6	658	4	US-10-441-949-44	Sequence 44, Appl
28	2432	16.5	657	3	US-09-851-859A-4	Sequence 4, Appl
29	2422	16.5	657	4	US-10-441-949-20	Sequence 20, Appl
30	2422	16.5	657	4	US-10-441-949-28	Sequence 28, Appl
31	2353	16.1	656	3	US-09-851-859A-3	Sequence 3, Appl
32	2007	13.7	471	4	US-10-448-871A-31	Sequence 31, Appl
33	1935	13.2	1438	5	US-10-450-763-42377	Sequence 42377, A
34	1867	12.8	471	4	US-10-448-871A-29	Sequence 29, Appl
35	1862	12.7	471	4	US-10-448-871A-27	Sequence 27, Appl
36	1850.5	12.6	470	4	US-10-448-871A-28	Sequence 28, Appl
37	1846	12.6	471	4	US-10-448-871A-30	Sequence 30, Appl
38	1821	12.4	1577	5	US-10-450-763-36834	Sequence 36834, A
39	1821	12.4	1577	5	US-10-450-763-57482	Sequence 57482, A
40	1784	12.2	473	4	US-10-448-871A-25	Sequence 25, Appl
41	1768.5	12.1	472	4	US-10-448-871A-26	Sequence 26, Appl
42	1572.5	10.7	1267	5	US-10-450-763-33781	Sequence 33781, A
43	1550.5	10.6	878	5	US-10-450-763-36843	Sequence 36843, A
44	1520	10.4	676	4	US-10-677-558-3	Sequence 3, Appl
45	1515	10.4	817	5	US-10-450-763-40498	Sequence 40498, A

ALIGNMENTS

RESULT 1

US-10-677-558-2  
; Sequence 2, Application US/10677558  
; Publication No. US20040096972A1  
; GENERAL INFORMATION:  
; APPLICANT: AUDIT, Muriel  
; APPLICANT: COSSET, Francois-Loic  
; TITLE OF INVENTION: CHIMERIC PLASMID COMPRISING A REPLICATIVE RETROVIRAL GENOME AND  
; FILE REFERENCE: 1759.135  
; CURRENT APPLICATION NUMBER: US/10/677,558  
; CURRENT FILING DATE: 2003-10-02  
; PRIOR APPLICATION NUMBER: PCT/FR02/03934  
; PRIOR FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: FR 0114976  
; PRIOR FILING DATE: 2001-11-20  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 1199  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: product of pol gene of pAM plasmid  
US-10-677-558-2

Alignment Scores:			
Pred. No.:	1.42e-256	Length:	1199
Score:	3980.50	Matches:	755
Percent Similarity:	76.2%	Conservative:	165
Best Local Similarity:	76.2%	Mismatches:	257
Query Match:	27.2%	Indels:	31
DB:	4	Gaps:	11

US-10-723-552-3 (1-8132) x US-10-677-558-2 (1-1199)



Db 720 ArgArgGlyLeuLeuThrSerGluGlyLysGluLeuLysAsnLysAspGluLeuLeuAla 739  
 QY 4371 CTATTAGAGCCGTACATTTTACCAAAAGGCTAGCTATTATATACACTGCTCTGACATCAG 4430  
 Db 740 LeuLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysProGlyHisGln 759  
 QY 4431 AAGACTAAGATCTCATATCCAGAGAAACCATGCTGACCGGGTTGCCAAGCAGGCA 4490  
 Db 760 LysGlyAsnSerAlaGluAlaArgGlyAsnArgMetAlaAspGlnAlaAlaArgGluVal 779  
 QY 4491 GCC-----CAGGGTGTAACTTCTCCCTATATAGAAATGCCAAGCCCA 4538  
 Db 780 AlaThrArgGluThrProGlyThrSerThrLeuLeuIleLysSerThrProTyrThr 799  
 QY 4539 GAACCCAGACGACAGTACACCTCAGAGAGCTGCAAGAGATAAAAGATAGACACAGTTC 4598  
 Db 800 HisGluHisPheHisThrValThrAspThrLysAspLeuThrLysLeu----- 816  
 QY 4599 TCTGAGACTCCGGAAGGAGCTGCTATACCTCA-----GATGGG 4637  
 Db 817 -----GlyAlaThrTyrAspSerAlaLysLysTyrTrpValTyrGlnGly 831  
 QY 4638 AAGAAATCTGCCCCACAAAGAGGTTAGATATGTCACAGATACATCGCTCAACC 4697  
 Db 832 LysProValMetProAspGlnPheThrPheGluLeuLeuAspPheLeuHisGlnLeuThr 851  
 QY 4698 CACCTAGGAACCTAAACACCTGCAGCAGTGGTGCAG-----ACATCCCTTATCATGTT 4751  
 Db 852 HisLeuSerPheSerLysThrLysAlaLeuLeuGluArgSerProSerProTyrTyrMet 871  
 QY 4752 CTGAGGCTACAGGAGTGTGCTGCTCGGTGGTCAAACTGTGTGCCCTGCCAGCTGGTT 4811  
 Db 872 LeuAsnArgAspArgThrLeuLysAsnIleThrGluThrCysLysAlaCysAlaGlnVal 891  
 QY 4812 RATGCTAATCTCCAGATGCTCCAGGAGAGACTAAGGGAAGCCACCCAGCGCT 4871  
 Db 892 AsnAlaSerLysSerAlaValLysGlnGlyThrArgValArgGlyHisArgProGlyThr 911  
 QY 4872 CACTGGGAAGTGCATCTACTAGGTAAAGCCGGCTAAATACGAAACAAATACCTATTG 4931  
 Db 912 HisTrpGluLeuAspPheThrGluValLysProGlyLeuTyrGlyTyrLysTyrLeuLeu 931  
 QY 4932 GTTTTGTAGACACCTTTTACAGATGGGTAGAGCTTATCTCTAAGAAAGAGACTTCA 4991  
 Db 932 ValPheValAspThrPheSerGlyTrpIleGluAlaPheProThrLysLysGluThrAla 951  
 QY 4992 ACGGTGGTGGTAAATACTGGAAGAAATTTTCCAGATTTGGAATACCTAAGTGA 5051  
 Db 952 LysValValThrLysLysLeuLeuGluGluIlePheProArgPheGlyMetProGlnVal 971  
 QY 5052 ATAGGTCAGACAAATCGTCCAGCTTTTGTGCCCAGGTAAAGTCAGGAGCTGGCCAAGATA 5111  
 Db 972 LeuGlyThrAspAsnGlyProAlaPheValSerLysValSerGlnThrValAlaAspLeu 991  
 QY 5112 TTGGGATGATTGGAAATCGATGTTGTCATACAGACCCCAAGCTCAGGACAGGTAGAG 5171  
 Db 992 LeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGlu 1011  
 QY 5172 AGGTGAATAGAACCAATTAAAGAGACCTTACTAATTAACCGCGAGACTGCGGTTAAT 5231  
 Db 1012 ArgMetAsnArgThrIleLysGluThrLeuThrLysLeuLeuAlaThrGlySerArg 1031  
 QY 5232 GATTGATAGCTCTCTGCGCTTTGTCCTTTTGGGTTAGGTTAGGAACACCCCTGCACAGTTT 5291  
 Db 1032 AspTrpValLeuLeuLeuProLeuAlaLeuTyrArgAlaArgAsnThrProGlyProHis 1051  
 QY 5292 GGGCTGACCCCTATGAATTACTCTACGGGGGACCCCGCCCAATGTGTAGAAATGCTTCT 5351  
 Db 1052 GlyLeuThrProTyrGluLeuLeuTyrGlyAlaProProProLeuValAsnPheProAsp 1071  
 QY 5352 GTACATAGTGTGAGCTGCTGCTTCCAGACCTTTGCTTCTAGGCTCAAGGACCTTAG 5411  
 Db 1072 ProAspMetThrArgValThrAsnSerProSerLeuGlnAlaHisLeuGlnAlaLeuTyr 1091

QY 5412 TGGGTGAGACAACGACGCTGGAGCAACTCCGGGAGGCCTACTCAGGAGGAGACTTG 5471  
 Db 1092 LeuValGlnHisGluValTrpArgProLeuAlaAlaAlaTyrGlnGluLeuAspArg 1111  
 QY 5472 CAG---ATCCACATCGTTTCCAGTGGGAGATTACGTCTACGTTAGAGCCACCGTCA 5528  
 Db 1112 ProValValProHisProTyrArgValGlyAspThrValTrpValArgHisGlnThr 1131  
 QY 5529 GAAACCTCGAGACTCGGTGGAGGCGCTTATCTCGTACTTTTGTGACCACCAACCGCT 5588  
 Db 1132 LysAsnLeuGluProArgTrpLysGlyProTyrThrValLeuLeuThrThrProThrAla 1151  
 QY 5589 GTGAAGTCAAGAGGAATCTCCACCTGGATCCATGCATCCACGCTTAAACCGGCGCCACCT 5648  
 Db 1152 LeuLysValAspGlyIleAlaAlaTrpIleHisAlaAlaHisValLysAlaAlaAspThr 1171  
 QY 5649 -----CCGATTCGGG-----TGAAGCGCAAGAACTGCAAAATCCCTTAAG 5693  
 Db 1172 GluSerGlyProSerSerGlyArgThrTrpArgValGlnArgSerGlnAsnProLeuLys 1191  
 QY 5694 CTTGCGCTCATCGCGTGGTTCCT 5717  
 Db 1192 IleArgLeuThrArgGlySerPro 1199

RESULT 2  
 ; US-09-851-859A-6  
 ; Sequence 6, Application US/09851859A  
 ; Patent No. US20020065407A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Banerjee, Papi T.  
 ; APPLICANT: Patience, Clive  
 ; APPLICANT: Andersson, Goxan K.  
 ; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of  
 ; TITLE OF INVENTION: Use  
 ; FILE REFERENCE: 61750-321  
 ; CURRENT APPLICATION NUMBER: US/09/851,859A  
 ; PRIOR FILING DATE: 1999-08-18  
 ; PRIOR APPLICATION NUMBER: 60/097,015  
 ; PRIOR FILING DATE: 1998-08-18  
 ; PRIOR APPLICATION NUMBER: US 09/376781  
 ; PRIOR FILING DATE: 2000-08-18  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 638  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:PERV-C  
 ; OTHER INFORMATION: polypeptide sequence taken from GenBank Accession  
 ; OTHER INFORMATION: No. US20020065407A1 AF038600 for comparison.  
 ; US-09-851-859A-6

Alignment Scores:  
 Pred. No.: 1,37e-219 Length: 638  
 Score: 3425.00 Matches: 637  
 Percent Similarity: 100.0% Conservative: 1  
 Best Local Similarity: 99.8% Mismatches: 0  
 Query Match: 23.4% Indels: 0  
 DB: 3 Gaps: 0

US-10-723-552-3 (1-8132) x US-09-851-859A-6 (1-638)

QY 5620 ATGCATCCACGCTTAAACCGGCGCCACCTCCCGATTCGGGTGGAAAGCCGAAAGACTG 5679  
 Db 1 MetHisProThrLeuAsnArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20  
 QY 5680 AAAATCCCTTAAGCTTCGCTCCATCGCTGCTTCTTACTCTGTCAATAACCTCTCAG 5739  
 Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
 QY 5740 ACTAATGGTATGCGCATAGGAGACGCTGAACTCCCAATAACCCCTTATCTCTCACTGG 5799



Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 5800 TTAATTAAGTACCTCCGACAGAGTATTAATATCAACAACACTCMAGGGAGGCTCCTTTA 5859  
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80  
QY 5860 GGAACCTGGTGGCTGATCATAGTTTGGCTCAGATCAGTATTCCTAGTCTGACCTCA 5919  
Db 81 GlyThrTrpProAspLeuThrValCysLeuArgSerValIleProSerLeuThrSer 100  
QY 5920 CCCCCAGATATCCTCCATGCTCAGGATTTTATGTTGCCAGGACCAACCAATAATGGA 5979  
Db 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120  
QY 5980 AAACATTGCGAAATCCAGAGATTTCTTTTGTAAACAATGGAAGTGTAACTCTAAT 6039  
Db 121 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 140  
QY 6040 GATGGATATTGGAATGCGCAACCTCTCAGCAGATAGGTAAAGTTTCTTATGTCAC 6099  
Db 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
QY 6100 ACCTATACCACTCTGGAACAATTTAATTACCTGACCTGGATTAGAATGGAAGCCCCAAG 6159  
Db 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180  
QY 6160 TGCTCTCTTCAGACCTAGATTACCTAAAAATAGTTTCTACTGAGAAAGAAAAACAAGAA 6219  
Db 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
QY 6220 AATATCTTAAATGGGTAAATGTTGCTCTGGGGAATGTTATATATGAGGCTCGGT 6279  
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrGlyGlySerGly 220  
QY 6280 AAACAACAGGCTCCATTCTAATCTATTCGCTCAAAATAAACAGCTGGAGCTCCAAATG 6339  
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
QY 6340 GCTATAGGACCAATACGGTCTTGACGGGTCAAGACCCCAACCCCAAGACGACGACCA 6399  
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
QY 6400 TCCTCTAACATACTCTGGATCAGACCCCACTGAGTCTAACAGCAGCAGCTAAATGGGG 6459  
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerSerSerThrThrLysMetGly 280  
QY 6460 GCAAACTTTTACGCTCATCCAGGGAGCTTTTCAAGCTCTTAACCTCCAGACTCCAGAG 6519  
Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 6520 GCTACCTCTTCTGTTGGCTATGCTTAGCTCGGGCCCACTTACTATGAAGGATGGCT 6579  
Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTrpGluGlyMetAla 320  
QY 6580 AGAAGGGGAAATCAATGTGACAAAGAACATAGACCAATGCACATGGGATCCCAA 6639  
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 6640 AATAAGCTTACCTTACTAGGTTTCTGGAAAAGGCACCTGCATAGGAAAGGTTTCCCCCA 6699  
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 6700 TCCCACCAACACCTTTGTAAACACACTGAAGCCTTTAATCAACCTCTGAGAGTCAATAT 6759  
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
QY 6760 CTGGTACCTGTTATGACAGTGTGGCATGTAATACTAGATTAAACCCCTTGTTTCC 6819  
Db 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 6820 ACCTTGGTTTTTAACCAAACTAAAGATTTTGGCATTATGGTCCAAATTTGTTCCCGAGTG 6879

Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 6880 TATTACTATCCGAAAAGCAATCCTTGTATGATATGACTACAGAAATCATCCACAAAAG 6939  
Db 421 TyrTyrTyrProGlnLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
QY 6940 AGAGAACCCATATCTCTGACACTTGCTCGACTTCTCGACTTGGAGTGGCAGCAGGTGTA 6999  
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaGlyVal 460  
QY 7000 GGAACAGGAACAGCTCCCTGGTCAAGGACCAACAGAGCTAGAAACAGGACTTAGTAAC 7059  
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 7060 CTATCATCGAATGCTAACAGAGATCTCCAGCCCTAGAAAACTCTCAGTAACCTGGAG 7119  
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 7120 GAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGAAAGAGGGTTAGATTATTATA 7179  
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
QY 7180 TTTCTAAAAGAGAGAGATTTATGTGTAGCTTGAAGAGGAATGCTGTTTTTATGTTGAT 7239  
Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAsp 540  
QY 7240 CATTCAAGGGCCATCAGAGACTCCATGMAACAAGCTTAGAGAAAGTTGGAGAACGTCGA 7299  
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArg 560  
QY 7300 AGGAAAAGGAAACTACTCAAGGGTGGTTTGAAGGATGGTTCAACAGGCTCTCTTGGTTG 7359  
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeu 580  
QY 7360 GCTACCTACTTTCTCTTTAAACAGACCTTAAATAGTCTCTCTCTGTTACTCACAGTT 7419  
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
QY 7420 GGGCCATGATTATTAAACAGTTAATTCCTTCAATTAGAGAACGAATAAGTCAGTCCAG 7479  
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 7480 ATCATGGTACTTAGACAAACAGTACCACCAAGCCCTCTAGCAGGGAAGCTGCCCGC 7533  
Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 638

RESULT 3  
US-10-441-949-32  
; Sequence 32, Application US/10441949  
; Publication No. US20040116684A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodrigo, Allen  
; APPLICANT: Ross, Howard A.  
; APPLICANT: Mullins, James I.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES  
; FILE REFERENCE: 08987-012001  
; CURRENT APPLICATION NUMBER: US/10/441, 949  
; PRIOR FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: PCT/US01/05288  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/183,659  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially generated peptide  
US-10-441-949-32  
Alignment Scores:

Pred. No.: 1.01e-218 Length: 638  
Score: 3412.00 Matches: 635  
Percent Similarity: 99.7% Conservative: 1  
Best Local Similarity: 99.5% Mismatches: 2  
Query Match: 23.3% Indels: 0  
DB: 4 Gaps: 0

US-10-723-552-3 (1-8132) x US-10-441-949-32 (1-638)

QY 5620 ATGCATCCCAAGTTAAACCGCGCCACCTCCGATTCGGGGTGAAGCGAAAGACTG 5679  
DB 1 MetHisProThrLeuSerArgArgHisLeuProIleArgGlyGlyProLysArgLeu 20  
QY 5680 AAAATCCCTTAAGCTTCGCTCATCGCTGGTTCCTTACTCTGCAATAAACCCTCTCAG 5739  
DB 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
QY 5740 ACTAATGGTATGGCATAGGACAGCCTCAACTCCCATAAACCTTATCTCCTCCTGG 5799  
DB 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 5800 TTAATTAAGTACCTCCGACAGCTATTATATCAACAACACTCAAGGGGAGGCTCCTTTA 5859  
DB 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyAlaProLeu 80  
QY 5860 GGAACCTGGTGGCTGATCTATACGTTTGCTCAGATCAGTTATCTAGTCTGACCTCA 5919  
DB 81 GlyThrTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
QY 5920 CCCCAGATATCTCCATGCTCAGGATTTTATGTTGGCCAGACCAACCAATATGGA 5979  
DB 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProAsnAsnGly 120  
QY 5980 AAACATTGCGAAATCCAGAGATTTCTTTGTAAACAATGGAACCTGTAACCTCTAAT 6039  
DB 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140  
QY 6040 GATGGATATGGAAATGGCAACCTCTCAGCAGGATAGGTAAGTTTCTTATGTCAAC 6099  
DB 141 AspGlyTyrTrpLysTrpProThrSerGlnAspArgValSerPheSerTyrValAsn 160  
QY 6100 ACTATACCACTCTCGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6159  
DB 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180  
QY 6160 TGCTCTCTCCATCAGACCTAGATTAACCTAAATAAAGTTTCACTGAGAAAGCAACAGAA 6219  
DB 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGlnLysGlyLysGlnGlu 200  
QY 6220 AATATCCTAAATGGGTAATGGTATGCTCTGGGGAATGGTATATTATGGAGCTCGGT 6279  
DB 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlySerGly 220  
QY 6280 AACACACCGCTCCATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6339  
DB 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProMet 240  
QY 6340 GCTATAGGACCAATACGGTCTTGACGGGTCAAGACCCCAACCAAGGACAGACCA 6399  
DB 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProThrGlnGlyProGlyPro 260  
QY 6400 TCCTCTAACATACTCTCGATCAGACCCCACTGAGCTTAACAGCAGCACTAAATGGGG 6459  
DB 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
QY 6460 GCMAAATCTTTAGCTCATCGGGAGCTTTTCAAGCTCTTAACTCCAGCTCCAGAG 6519  
DB 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 6520 GCTACCTCTCTTGTGGCTATGCTTAGCTTCGGGCCCACTTACTATCAAGGAATGGCT 6579  
DB 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAla 320

QY 6580 AGAAGAGGGAATTTCAATGTGACAAAGAACATAGAGACCAATGCATCGGGATCCCAA 6639  
DB ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 6640 AATAAGCTTACCTTACTGAGGTTTCTGGAAGAGGACCTGCATAGAAAGGTTCCCCCA 6699  
DB AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 6700 TCCCACCAACACCTTTGTAACCACTGAAGCCCTTAAATCAAACTCTGAGAGTCAATAT 6759  
DB SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
QY 6760 CTGGTACCTGGTATGACAGGTGGTGGCATGTAATATCTGATTAACCCCTTGTGTTC 6819  
DB LeuValProGlyTyrAspArgTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 6820 ACCTGGTGTTTTAAACCAAACTAAAGATTTTGCATTTATGGTCCAAATTTGTTCCCGAGTG 6879  
DB ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 6880 TATTACTATCCCAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAG 6939  
DB TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
QY 6940 AGAAGACCATATCTCTGACACTTGTCTGCTCGGACTTGGAGTGGCAGCAGGTGA 6999  
DB ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaGlyVal 460  
QY 7000 GGAACGAGAACAGCTGCTGTCAGGGACCAACAGCAGCTAGAAAACAGACTTAGTAAC 7059  
DB GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 7060 CTACATCGAATTTGTAACAGAGATCTCCAAGCCCTAGAAAAATCTCTAGTAACCTGGAG 7119  
DB LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 7120 GAATCCCTAACCTCTTATCTGAAGTAGTCTACAGATAGAGAGGGTTAGATTATTATTA 7179  
DB GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
QY 7180 TTTCTAAAGAGAGGAGGATTTATGTAGCTTGAAGGAGGAATGCTGTTTATGTGGAT 7239  
DB PheLeuLysGluGlyLeuCysValAlaLeuLysGluCysCysPheTyrValAsp 540  
QY 7240 CATTCAGGGCCCATCAGACACTCCATGAACAAGCTTAGAGAAAGGTTGAGAGAGCTCGA 7299  
DB HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuLysArgArg 560  
QY 7300 AGGGAAGGAAACTACTCAAGGGTGGTTTGAGGGATGGTTCAACAGGCTCTCTTTGGTTG 7359  
DB ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580  
QY 7360 GCTACCTCTTCTGCTTTAAACAGGACCTTAATAGTCTCTCTCTCTGTTACTCAGATT 7419  
DB AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
QY 7420 GGGCCATGTATTATTAAACAAGTTAATGCTTCATTAGAGAACGATTAAGTCAGTCCAG 7479  
DB GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 7480 ATCATGGTACTTAGACCAACAGTACCAGCCCGCTCTAGCAGGGAAGCTGGCCGC 7533  
DB IleMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 638

## RESULT 4

US-10-441-949-34  
; Sequence 34, Application US/10441949  
; Publication No. US20040116684A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodrigo, Allen  
; APPLICANT: Ross, Howard A.  
; APPLICANT: Mullins, James I.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES

```
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-34

Alignment Scores:
Pred. No.: 1,01e-218 Length: 638
Score: 3412.00 Matches: 635
Percent Similarity: 99.7% Conservative: 1
Best Local Similarity: 99.5% Mismatches: 2
Query Match: 23.3% Indels: 0
DB: 4 Gaps: 0

US-10-723-552-3 (1-8132) x US-10-441-949-34 (1-638)

QY 5620 ATGCATCCACGGTAAACCGCGCCACCTCCGANTCGGGGTGGAAGCGAAAGACTG 5679
DB 1 MetHisProThrLeuSerArgHisLeuProileArgGlyLysProLysArgLeu 20
QY 5680 ARAATCCCTTAGCTCGCTCCATCGCTGGTTCCTTACTCTGCTCAATAAATCTCAG 5739
DB 21 LysileProLeuSerPheAlaSerileAlaTrpPheLeuThrLeuSerileThrSerGln 40
QY 5740 ACTAATGGTATGGCATAGGAGACAGCTCACTCCCATAAACCTTATCTCACCTGG 5799
DB 41 ThrAsnGlyMetArgileGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 5800 TTAATTAAGTCCGCGCAGAGTATTAATATCAACAACTACAGGGAGGCTCCTTTA 5859
DB 61 LeuileThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80
QY 5860 GGAACCTGGTGGCTGATATACGTTTGGCTCAGATCAGTATTCTAGTCTGACCTCA 5919
DB 81 GlyThrTrpProAspLeuTyValCysLeuArgSerValIleProSerLeuThrSer 100
QY 5920 CCCCAGATATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACCAACCAATAATGGA 5979
DB 101 ProProAspileLeuHisAlaHisGlyPheTyValCysProGlyProProAsnAsnGly 120
QY 5980 AAACATTCGGGAATCCAGAGATTTCTTTTGTAAACATGGAAGTGTAACTCTAAT 6039
DB 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140
QY 6040 GATGGATATTGGAATCGCAACTCTCAGCAGATAGGTAGTATTTCTTATGTCAC 6099
DB 141 AspGlyTyTrpLysTrpProThrSerGlnAspArgValSerPheSerTyValAsn 160
QY 6100 ACCTATACCAAGCTCTGACAAATTTAAATTAACCTGACCTGGATTTAGAACTGGAAGCCCCAAG 6159
DB 161 ThrTyThrSerSerGlyGlnPheAsnTyLeuThrTrpIleArgThrGlySerProLys 180
QY 6160 TGCTCTCCTCAGACCTAGATTACTAAATAAGTTTCTACTGAGAAGGAAACAAGAA 6219
DB 181 CysSerProSerAspLeuAspTyLeuLysileSerPheThrGluLysGlyLysGlnGlu 200
QY 6220 AATATCTAAATCGGTAAATGTTATCTCTGGGAATGGTATATTATGAGGCTCGGT 6279
DB 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyTrpGlyLysGly 220
QY 6280 AAACAACGAGCTCCATTCTAACTATTTCGCTCAAAATAAACACGCTGGAGCCTCCCAATG 6339
DB 1
```

```
DB 221 LysGlnProGlySerileLeuThrIleArgLeuLysileAsnGlnLeuGluProProMet 240
QY 6340 GCTATAGGACCAATAACGGTCTTGACGGGTCAAGACCCCAACCAAGCAGGACCA 6399
DB 241 AlaileGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260
QY 6400 TCCTCTAAACATTAATCTCTGATCAGACCCCACTGAGTCTTAACAGCAGCAGCTAAATGGGG 6459
DB 261 SerSerAsnileThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280
QY 6460 GCAAAACTTTTAGCCCTATCCAGGAGCTTTCAAGCTTTAACTCCAGACTCCAGAG 6519
DB 281 AlaLysLeuPheSerLeuileGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300
QY 6520 GCTACCTCTTCTGTGGCTATCTTAGCTTCGGGCCCACTTACTATGAAGAAATGGCT 6579
DB 301 AlaThrSerSerCysTrpLeuLysLeuAlaLeuGlyProProTyTrpGlyGlyMetAla 320
QY 6580 AGAAGAGGGAAATTCATGTGACAAAGAACATAGAGACCAATGCACATGGGGATCCCA 6639
DB 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340
QY 6640 AATAAGCTTACCTTACTGAGGTTTCTGGAAGAGGACCTGCTAGATAGAAAGGTTCCCCCA 6699
DB 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysileGlyLysValProPro 360
QY 6700 TCCCACCAACACCTTTGTACCACTGAGCCTTAAATCAACCTCTGAGAGTCAATAT 6759
DB 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTy 380
QY 6760 CTGGTACCTGGTTATGACAGGCTGGTGGCATGTAAATCTGGATTAACCCCTTGTGTTC 6819
DB 381 LeuValProGlyTyTrpAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400
QY 6820 ACCTTGGTTTTTAAACCAACTAAGATTTTTCATTTATGCTCAAAATTTTCCCGAGTG 6879
DB 401 ThrLeuValPheAsnGlnThrLysAspPheCysileMetValGlnIleValProArgVal 420
QY 6880 TATTACTATCCCAAAAGCAATCCTTGATGATATGACTACAGAAATCATCCACAAAG 6939
DB 421 TyTyTyTrpProGlnLysAlaIleLeuAspGlyTyTrpAspTyTrpArgAsnHisArgGlnLys 440
QY 6940 AGAAGAACCAATCTCTGCACACTTGTCTGATGCTCGACTTGGAGTGGCAGCAGGTGTA 6999
DB 441 ArgGluProileSerLeuThrLeuAlaValIleLeuGlyLeuGlyValAlaAlaGlyVal 460
QY 7000 GGAACAGGAACACGCTGCCCTGCTCAGGACCAACAGCAGCTAGAACAGCAGCTAGTAAC 7059
DB 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
QY 7060 CTACATCGAATTCGTAACAGAGATCTCCAAGCCCTAGAAAAATCTGTCACTAACCTGGAG 7119
DB 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGlnLysSerValSerAsnLeuGlu 500
QY 7120 GAATCCCTAACTCTTATCTGAAGTAGTCTCCTACAGAAATAGAAGGGTTAGATTTATTA 7179
DB 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520
QY 7180 TTTCTAAAGAGAGAGATTATGTGTAGCTTGAAGAGGAATGCTCTTTTATGTGGAT 7239
DB 521 PheLeuLysGluGlyLeuCysValAlaLeuLysGluCysCysPheTyValAsp 540
QY 7240 CATTTCGGGCCCATCAGAGACTCCATGAACAGCTTAGAAAGGTTGGAGAGCGTCGA 7299
DB 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560
QY 7300 AGGAAAAAGGAAACTACTCAAGGGTGGTTTGGAGGATGGTTTCAACAGGCTCTCTTGGTTG 7359
DB 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580
QY 7360 GCTACCCCTACTTCTGCTTTAAACAGGACCCCTTAATAGTCTCTCTCTCTTACTCAGATT 7419
DB 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600
```

QY 7420 GGGCCATGTTATTAAACAAGTTAAATTCCTTCATTAGAGAACGAATAAGTGCAGTCCAG 7479  
Db |||||  
601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 7480 ATCATGCTACTAGACACAGTACCAAGCCCGTCTAGCAGGGAAGCTGGCCG 7533  
Db |||||  
621 IleMetValLeuArgGlnGlnIleArgGlnSerProSerArgGluAlaGlyArg 638  
RESULT 5  
US-10-441-949-40  
; Sequence 40. Application US/10441949  
; Publication No. US20040116684A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodrigo, Allen  
; APPLICANT: Ross, Howard A.  
; APPLICANT: Mullins, James I.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES  
; FILE REFERENCE: 08987-012001  
; CURRENT APPLICATION NUMBER: US/10/441,949  
; PRIOR FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: PCT/US01/05288  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/183,659  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 40  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially generated peptide  
US-10-441-949-40  
Alignment Scores:  
Pred. No.: 1.01e-218 Length: 638  
Score: 3412.00 Matches: 635  
Percent Similarity: 99.7% Conservative: 1  
Best Local Similarity: 99.5% Mismatches: 2  
Query Match: 23.3% Indels: 0  
DB: 4 Gaps: 0  
US-10-723-552-3 (1-8132) x US-10-441-949-40 (1-638)  
QY 5620 ATGCATCCCAAGTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAGCGAAAGACTG 5679  
Db |||||  
1 MethIleProThrLeuSerArgGlnHisLeuProIleArgGlyLysProLysArgLeu 20  
QY 5680 AAAATCCCTTAAGCTTCGCTCCATCGCGTGTTCCTTACTCTGTCAATAAACCCTCTCAG 5739  
Db |||||  
21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
QY 5740 ACTAATGGTATGGCATAGAGACAGCTGAACCTCCCAATAAACCTTATCTCACTGG 5799  
Db |||||  
41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 5800 TTAATTACTACCTCCGCGACAGGTATTAAATCAACAACACTCAAGGGGAGGCTCCTTTA 5859  
Db |||||  
61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyAlaProLeu 80  
QY 5860 GGAACCTGGTGGCTGATCTATATACGTTGCTCAGATCAGTTATTCCTAGTCTGACCTCA 5919  
Db |||||  
81 GlyThrTrpProAspLeuTrpValCysLeuArgSerValIleProSerLeuThrSer 100  
QY 5920 CCCCCAGATATCTCATGCTACGGATTTTATGTTGGCCAGGACACCAATAATAGGA 5979  
Db |||||  
101 ProProAspIleLeuHisAlaHisGlyPheTyValCysProGlyProProAsnAsnGly 120  
QY 5980 AAACATTGGGAATCCCGAGATTTCTTTTGTAAACATGGAACGTGTAACTCTAAT 6039  
Db |||||  
121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140

QY 6040 GATGGATATTGGAAATGGCCAACCTCTCAGCAGGATAGGTAACTTTTCTTATGTCAAC 6099  
Db |||||  
141 AspGlyTyrrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrrValAsn 160  
QY 6100 ACCTATACGAGCTCTGCACAAATTAAATTACTGACCTGGATTAGAACTGGAGCCCCAAG 6159  
Db |||||  
161 ThrTyrrThrSerSerGlyGlnPheAsnTyrrLeuThrTrpIleArgThrGlySerProLys 180  
QY 6160 TGCTCTCTCTCAGACCTAGATTACTAAAAATAGTTTCTACTGAGAAGGAAAAACAAG 6219  
Db |||||  
181 CysSerProSerAspLeuAspTyrrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
QY 6220 AATATCTAAAAATGGGTAAATGGTATGCTCTTGGGGAATGGTATATATATGAGGCTCGGGT 6279  
Db |||||  
201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrrGlyGlySerGly 220  
QY 6280 AAACACACGAGCTCTCTTAACTATTTCGCTCAAAATAAACCCAGCTGGAGCTCCCATG 6339  
Db |||||  
221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
QY 6340 GCTATAGGACCAATACGGTCTTACGGGTCAAAAGACCCCAACCCCAAGGACACGACCA 6399  
Db |||||  
241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
QY 6400 TCCTCTAACATAACTTCTGGATCAGACCCCACTGAGCTAACACGACGACTAAAAATGGG 6459  
Db |||||  
261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
QY 6460 GCAAACTTTTAGCCTCATCAGGAGCTTTTCAAGCTCTTAACTCCAGCTCCAGAG 6519  
Db |||||  
281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 6520 GCTACCTCTTCTTGGTCTATCCTTAGCTTCGGGCCACCTTACTATGAGGAATGGCT 6579  
Db |||||  
301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrrTyrrGluGlyMecAla 320  
QY 6580 AGAAGAGGGAATTCATATGTGACAAAAGAACATAGAGACCAATGCATGGGATCCCAA 6639  
Db |||||  
321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 6640 AATAAGCTTACCCTTACTAGGTTCTTGGAAAGGACCTGCTAGGAAAGTTCCCCCA 6699  
Db |||||  
341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 6700 TCCCAACCAACACTTGTAAACCACTGAAGCCTTAAATCAAACTCTGAGAGTCAATAT 6759  
Db |||||  
361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyrr 380  
QY 6760 CTGGTACCTGGTTATGACAGGTGGTGGCATGTAATACCTGATTAACCCCTTGTGTTTC 6819  
Db |||||  
381 LeuValProGlyTyrrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 6820 ACCTTGGTTTTAACCAAACTAAAGATTTTGTGATTATGGTCCAAATGTTCCCCGAGTG 6879  
Db |||||  
401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 6880 TATTACTATCCCGAAAAAGCAATCCTTGATGATATGACTACAGAAATCATCCACAAAG 6939  
Db |||||  
421 TyrrTyrrProGluLysAlaIleLeuAspGluTyrrAspTyrrArgAsnHisArgGlnLys 440  
QY 6940 AGAAGAACCATATCTCTGACACTTGTGATGCTCGGACTTGGAGTGGCAGGAGGTGA 6999  
Db |||||  
441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
QY 7000 GGAACGAGAACAGCTGCCCTGTGTCAGGACACACAGCAGCTAGAAAACAGGACTTAGTAAC 7059  
Db |||||  
461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 7060 CTACATCGAATTTGTAACAGAAGTCTCCAAAGCCCTAGAAAATCTCTCAGTACCTGGAG 7119  
Db |||||  
481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 7120 GAATCCCTAACCTCTTATCTGAAGTAGTCTTACAGAATAGAGAGGGTTAGATTATTA 7179



QY 6880 TATTACTATCCCGAAAGCAATCTTGTATGAATATGACTACAGAAATCATGCAAAAAG 6939  
Db 421 Tyr-Tyr-Tyr-ProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
QY 6940 AGAGAACCCATATCTGACACTTGTGTGATCTGGACTTGGAGTGGCAGAGGTGTA 6999  
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
QY 7000 GGAACAGGAACAGCTGCCCTGGTGCACGGGACACAGCAGCTAGAAAACAGGACTTAGTAAC 7059  
Db 461 GlyThrGlyThrAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 7060 CTACATCGAATTGTAAACAGAGATCTCAAGCCCTAGAAAAATCTGTCACTACCTGGAG 7119  
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 7120 GAATCCCTACCTCCTTATCTGAGTAGTCTCAGAGATAGAGAGAGGTTATTATTA 7179  
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
QY 7180 TTTCTAAAAGAGAGGATTTATGTAGCTTCAAGGAGGAATCTGTTTTTATGTGGAT 7239  
Db 521 PheLeuLysGluGlyLeuCysValAlaLeuLysGluGluCysPheTyrValAsp 540  
QY 7240 CATTACAGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGTTGGAGAGCGTCGA 7299  
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560  
QY 7300 AGGAAAGAACTACTCAAGGGTGTGGAGGATGGTTCAACAGGCTCTTTGGTGG 7359  
Db 561 ArgGluLysGluThrThrGlnGlyTyrPheGluGlyTyrPheAsnArgSerProTyrLeu 580  
QY 7360 GCTACCTCTCTCTCTTAAACAGACCTTAATAGTCTCTCTCTCTCTCTCTCTCT 7419  
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrVal 600  
QY 7420 GGCCCATGTATTATTAACAGTTAATGCTTCATTAGAGAACGAATAAGTGCAGTCCAG 7479  
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 7480 ATCATGGTACTAGACACAGTACCAAGCCCTCTAGCAGGAGCTGGCCGC 7533  
Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 638

RESULT 7

US-10-441-949-45  
; Sequence 45, Application US/10441949  
; Publication No. US20040116694A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodrigo, Allen  
; APPLICANT: Rosa, Howard A.  
; APPLICANT: Mullins, James I.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES  
; FILE REFERENCE: 08987-012001  
; CURRENT APPLICATION NUMBER: US/10/441,949  
; PRIOR FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: PCT/US01/05288  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/183,659  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially generated peptide  
US-10-441-949-45

Alignment Scores:

Pred. No.: 1.01e-218 Length: 638

Score: 3412.00 Matches: 635  
Percent Similarity: 99.7% Conservative: 1  
Best Local Similarity: 99.5% Mismatches: 2  
Query Match: 23.3% Indels: 0  
DB: 4 Gaps: 0  
US-10-723-552-3 (1-8132) x US-10-441-949-45 (1-638)  
QY 5620 ATGCATCCCAACGTTAAACCGCGCCACCTCCCGATTTCGGGTGAAAGCCGAAAGACTG 5679  
Db 1 MetHisProThrLeuSerArgArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20  
QY 5680 AAAATCCCTTAAGCTTCGCTCCATCGCGTGGTCTCTTACTCTGTCAATAAACCCTCTCAG 5739  
Db 21 LysIleProLeuSerPheAlaSerIleAlaTyrPheLeuThrLeuSerIleThrSerGln 40  
QY 5740 ACTAATGGTATCGGCATAGGAGACAGCTCACTCCATAAAACCCCTATCTCTCACCCTGG 5799  
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTyr 60  
QY 5800 TTAATTACTGACTCCCGCACAGGTATTAATATCAACAACAACACTCAAGGGGAGGCTCCTTTA 5859  
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80  
QY 5860 GGAACCTGTGGCCTGATCTATACGTTTCCTCAGATCAGTTATCTAGTCTGACCTCA 5919  
Db 81 GlyThrTyrTyrProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
QY 5920 CCCCAGATATCTTCATGCTCAGGATTTTATGTTTGGCCAGGACACCAATAATATGGA 5979  
Db 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120  
QY 5980 AAACATTGGGAAATCCAGAGATTTCTTTGTAAACAATGGAACCTGTGAACCTCTAAT 6039  
Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTyrAsnCysValThrSerAsn 140  
QY 6040 GATGATATTGGAAATGGCAACCTCTCAGCAGGATAGGTTAGTTTCTTATGATCAAC 6099  
Db 141 AspGlyTyrTyrLysTyrProThrSerGlnAspArgValSerPheSerTyrValAsn 160  
QY 6100 ACTATACAGCTCTGGACAATTAATTAACCTGACCTGATAGACTAGACTGGAAGCCCAAG 6159  
Db 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTyrIleArgThrGlySerProLys 180  
QY 6160 TGTCTCTCTTCAGACCTAGATTACCTTAAATAAAGTTTTCACCTGAGAAAGGAAACAGAA 6219  
Db 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
QY 6220 AATATCCTAAAAATGGTAAATGGTATGCTCTGGGGAATGGTATATTATGAGGCTCGGGT 6279  
Db 201 AsnIleLeuLysTyrValAsnGlyMetSerTyrGlyMetValTyrTyrGlySerGly 220  
QY 6280 AAACAACAGGCTCAATTTAACTATTTCGCTCAAAATAAACCAGCTGAGCTTCCAATG 6339  
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
QY 6340 GCTATAGGACCAATAACGCTTTCAGGCTCAAGACCCCAACCCCAAGACAGACCA 6399  
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
QY 6400 TCCTCTAACATAACTTCTGGATCAGACCCCACTGAGTCTTAACAGACAGCACTAAAAATGGG 6459  
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
QY 6460 GCAAACTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCAGCTCCAGAG 6519  
Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 6520 GCTACTCTCTTGTGTGGCTATGCTAGCTTCGGGCCCACTTACTATCAAGGAATGGCT 6579  
Db 301 AlaThrSerSerCysTyrLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAla 320  
QY 6580 AGAAGAGGGAATTCATGTGCAAAAAAGAAACATAGAGACCAATGCACATGGGATCCCAA 6639

```
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340
QY 6640 AATAAGCTTACCTTACTAGGTTTCTGGAAAGGCACCTGCATAGGAAAGTTTCCCCCA 6699
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360
QY 6700 TCCCACCAACACCTTTGTACCAACACTGAAGCCTTTAATCAACCTCTGAGAGTCAATAT 6759
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380
QY 6760 CTGCTACTCTGTTATGACAGCTGGTGGCCATGTAAATACTCGATTAAACCCCTGTGTTTCC 6819
Db 381 LeuValProGlyTyrAspAsgTTPAlaCysAsnThrGlyLeuThrProCysValSer 400
QY 6820 ACCTTGTTTTTAAACCAACTAAAGATTTTGCATTATGCTCCAAATTTGTTCCCGAGTG 6879
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420
QY 6880 TATTACTATCCGAAAAGCAATCCTTGATGATATATGACTACAGAAATCATCGACAAAAG 6939
Db 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisAsgGlnLys 440
QY 6940 AGAAGACCCATATCTCTGACACTTGTCTGATGCTCGGACTTGGAGTGGCAGCAGGTGTA 6999
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIleGlyVal 460
QY 7000 GGAACAGGAACAGTCCCTGGTCCAGGACACAGCAGCTAGAAAACAGGACTTAGTAAC 7059
Db 461 GlyThrGlyThrAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
QY 7060 CTACATCGAATTGTAACAGAAGATCTCAAGCCCTAGAAAATCTGTCACTAACCTGGAG 7119
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500
QY 7120 GAATCCCTAACCTCTTATCTGAAGTAGTCTCAGAAATAGAGAGGTTAGATTATTA 7179
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520
QY 7180 TTTCTAAAAGAGGAGATTATGTGTAGCTTGAAGGAGAAATGCTGTTTTTAATGTGAT 7239
Db 521 PheLeuLysGluGlyLysCysValAlaLeuLysGluGlyCysCysPheTyrValAsp 540
QY 7240 CATTGAGGGCCATCAGAGCTCCATGAACAAGCTTAGAGAAAGTTGGAGAGCCTCGA 7299
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysAsgArg 560
QY 7300 AGGMAAAGGAACTACTCAAGGTGGTTTGGAGGATGGTTCAACAGGTCTCTTTGGTTG 7359
Db 561 ArgGluLysGluThrThrGlnGlyTyrPheGluGlyTyrPheAsnArgSerProTrpLeu 580
QY 7360 GCTACCCCTACTTTCTGTTTAAACAGGACCTTAAATAGTCTCTCTCTGTTTACTACAGTT 7419
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrVal 600
QY 7420 GGCCCATGTTATTAAACAAGTTAATTGCTTCAATAGAGAACGAATAAGTGCAGTCCAG 7479
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620
QY 7480 ATCATGTTACTTAGACAAACAGTACCAAGCCGCTAGCAGGGAAGCTGGCCGC 7533
Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyAsg 638
```

## RESULT 8

```
US-10-441-949-36
; Sequence 36, Application US/10441949
; Publication No. US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodrigo, Allen
; APPLICANT: Ross, Howard A.
; APPLICANT: Mullins, James I.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
; FILE REFERENCE: 08987-012001
```

```
; CURRENT APPLICATION NUMBER: US/10/441,949
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-36
```

```
Alignment Scores:
Pred. No.: 1,8e-212 Length: 678
Score: 3318.50 Matches: 636
Percent Similarity: 93.8% Conservative: 1
Best Local Similarity: 93.7% Mismatches: 3
Query Match: 22.7% Indels: 39
DB: 4 Gaps: 8
```

US-10-723-552-3 (1-8132) x US-10-441-949-36 (1-678)

```
QY 5620 ATGCATCCACAGTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAGCCGAAAGACTG 5679
Db 1 MetHisProThrLeuSerArgHisLeuProIleArgGlyLysProLysArgLeu 20
QY 5680 ARAATCCCTTAGCTTCGCTCCATCGCGTGTCTTACTCTGTCATCAATAACTCTCAG 5739
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40
QY 5740 ACTAATCGGTATGCGCATAGAGACAGCCCTGAACTCCCATAAACCTTATCTCACCTGG 5799
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 5800 TTAATTACTGACTCCGGCAGAGTATATATACACACTCAAGGGGAGGCTCTTTTA 5859
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80
QY 5860 GGAACCTGGTGGCTCATCTATACGTTTGCTCAGATCAGTATTCTTCTAGTCTG----- 5913
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuAsnAsp 100
QY 5914 ---ACCTCACCCCCAGATATCTCCATGCTCAGGATTTTATGTTGCCCAGGACCA 5970
Db 101 GlnThrSerProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProPro 120
QY 5971 AATAATGGAACAATTGCGGAATCCAGAGATTTCTTTTGTAAACAATGGAAGTGTGTA 6030
Db 121 AsnAsnGlyLysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysVal 140
QY 6031 ACCTCTAATGATGATATTGGAATGCGCAACCTCTCAGCAGGATAGGGTAAGTTTCT 6090
Db 141 ThrSerAsnAspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSer 160
QY 6091 TATGTCAACACCTTATACCACTCTGGCAATTTAATTATC-----CTGACC 6135
Db 161 TyrValAsnThrTyrThrSerSerGlyGlnPheAsnTyrGlyHisGlyAsgTrpLeuThr 180
QY 6136 TGG-----ATTAGACTGGAAGCCCCAAGTGCTCTCTCTCAGAC 6174
Db 181 TrpGlnGlnArgValGlnLysAspIleArgThrGlySerProLysCysSerProSerAsp 200
QY 6175 CTAGATTACCTAAATAAGTTTCACTGAGAAAGGAAACAAAGAAATATCTCTAAAATGG 6234
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleLeuLysTrp 220
QY 6235 GTAAATGGTATGCTTTGGGGAATGGTATATATTGAGAGGCTCGGGTAAACACAGGCTCC 6294
Db 221 ValAsnGlyMetSerTrpGlyMetValTyrTyrGlySerGlyLysGlnProGlySer 240
```



```
QY 6295 ATTCTAATATTGCGCTCAAAATAAAC---CAGCTGGAGGCTTCAATGGCTATAGGACCA 6351
Db 241 ILELeuThrIleArgLeuLysIleAsnThrGlnLeuGluProProMetAlaIleGlyPro 260
QY 6352 AATACGGCTCTGACGGTCAAGACCCCAACCCAGGACCA-----6393
Db 261 AsnThrValLeuThrGlyGlnArgProProThrGlnGlyProProHisAsnLeuProVal 280
QY 6394 -----GGACCATCCTCT-----AACATAACTTCTGGA 6420
Db 281 ProGlnGlyProSerProAsnProAspIleThrGlnSerAspTyrAsnIleThrSerGly 300
QY 6421 TCAGACCCCACT-----GAGTCTAACAGCAGCACTAAATATGGGGCAAAA 6465
Db 301 SerAspProThrAsnThrProArgAsnGluSerAsnSerThrThrLysMetGlyAlaLys 320
QY 6466 CTTTATAGCTCATCCAGGAGCTTTCAAGCTCTTAAGTCCAGGACTCCAGAGGCTACC 6525
Db 321 LeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThr 340
QY 6526 TCTTCTTGTGTGCTATGCTTCCGGGCCACCTTACTATGAAGGAATGGCTAGAAGA 6585
Db 341 SerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTrpGlyMetAlaArgArg 360
QY 6586 GGGAAATTCATGTGACAAAAGAACATAGAGACCAATGCATGGGATCCCAAAATAAG 6645
Db 361 GlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGlnAsnLys 380
QY 6646 CTTACCTTACTCAGGTTCTCGAAAAGGCACCTGCATAGGAAGTTCCTCCCATCCAC 6705
Db 381 LeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProProSerHis 400
QY 6706 CAACACCTTGTGAACACACTGAAGCTTTAATAACAACTCTCAGAGTCAATATCGTA 6765
Db 401 GlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyrLeuVal 420
QY 6766 CTTGGTTATGACAGGTGGTGGGCGATGTAATPACTGGATTAAACCCCTTGTGTTCCACCTTG 6825
Db 421 ProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeu 440
QY 6826 GTTTTAAACCAACTAAGATTTTTCATTATGTCGAATTCGTCGAGTGTATTAC 6885
Db 441 ValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrTrp 460
QY 6886 TATCCCGAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAAGAGAGAA 6945
Db 461 TyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLysArgGlu 480
QY 6946 CCATATCTCTGACACTTGTGTGATGCTCGGACTTGGAGTGGCAGAGTGTAGGAACA 7005
Db 481 ProfileSerLeuThrLeuAlaValMetLeuLeuGlyValAlaAlaGlyValGlyThr 500
QY 7006 GGAACAGTCCCTGTGTCAGGACACACAGCAGCTAGAACAGGACTTAGTACCTACAT 7065
Db 501 GlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHis 520
QY 7066 CGAATTGTAAACAGAAGATCTCCAAGCCCTAGAAAATCTGCAGTAACCTGGAGGAATCC 7125
Db 521 ArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluLysSer 540
QY 7126 CTAACCTCTCTTCTGAAGTAGTCTTACAGAATAGAGAGGGTGTAGATTTATTTCCTA 7185
Db 541 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu 560
QY 7186 AAGAAGAGGAGTATTGTGTAGCTTGAAGGAGGAATGCTGTTTTTATGTGATCATTTCA 7245
Db 561 LysGluGlyGlyLeuCysValAlaLeuLysGluLysCysPheTyrValAspHisSer 580
QY 7246 GGGGCCATCAGACATCCATGAACAGCTTAGAGAAAGTTCGAGAGCGTCCAGGGAA 7305
Db 581 GlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArgGlu 600
```

```
QY 7306 AAGAAACTACTCAAGGTTGTTTGAGGATGTTCAACAGGTTCTTTTGGTTGGCTACC 7365
Db 601 LysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeuAlaThr 620
QY 7366 CTACTTCTGCTTTAACAGGACCCCTTAATAGTCTCTCTCTCTCTCTCAGTTGGGCCA 7425
Db 621 LeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrValGlyPro 640
QY 7426 TGTATTATTAAACAAGTTAATTGCTTCTATTAGAGAAGTAATAAGTGCAGTCCAGATCATG 7485
Db 641 CysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMet 660
QY 7486 GTACTTAGACAAACAGTACCACCAAGCCCTCTAGCAGGAAAGCTGCCGCTAGCTCTAC 7542
Db 661 ValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg---LeuTyr 678

RESULT 9
US-10-441-949-38
; Sequence 38, Application US/10441949
; Publication No. US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodrigo, Allen
; APPLICANT: Ross, Howard A.
; APPLICANT: Mullins, James I.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-38

Alignment Scores:
Pred. No.: 1,8e-212 Length: 678
Score: 3318.50 Matches: 636
Percent Similarity: 93.8% Conservative: 1
Best Local Similarity: 93.7% Mismatches: 3
Query Match: 22.7% Indels: 39
DB: 4 Gaps: 8

US-10-723-552-3 (1-8132) x US-10-441-949-38 (1-678)

QY 5620 ATGCATCCACAGCTTAAACCGGCGCCACCTCCCGATTCCGGGTGGAAAGCGGAAAAGACTG 5679
Db 1 MethisProThrLeuSerArgArgHisLeuProIleArgGlyLysProLysArgLeu 20
QY 5680 AAAATCCCTTAAAGCTTGCCTCCATCGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTCAG 5739
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40
QY 5740 ACTAATGGTATGCGCATAGGACAGCCTGAAGTCTCCATAAACCTTATCTCTCACCTGG 5799
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 5800 TTAATTAAGTCTCCGGCACAGGTATTATATATATATATATATATATATATATATATATAT 5859
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80
QY 5860 GGAACTGTGTGCTGATCTATACATGTTGCTTGCCTCAGATCAGTTATTCCTAGTCTG --- 5913
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuA-gserValIleProSerLeuAsnAep 100
QY 5914 ---ACCTCAACCCCGCATATCTCCATGCTCAGGATTTTATGTTTGGCCGACGACCACCA 5970
```

```
Db      101  GlnThrSerProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProPro 120
Qy      5971  AATAATGGAAACAAATTCGGGAATCCAGAGATTTCTTTTGTAAACAATGGAACTGTGTA 6030
Db      121  AsnAsnGlyLysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysVal 140
Qy      6031  ACCTCTAATCATGATATTCGAATGCGCAATGCGCACTCTCAGCAGATAGGTAAGTTTCT 6090
Db      141  ThrSerAsnAspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSer 160
Qy      6091  TATGTCAACCACTATACCACTCTGGACAATTTAATATAC-----CTGACC 6135
Db      161  TyrValAsnThrTyrThrSerSerGlyGlnPheAsnTyrGlyHisGlyArgTrpLeuThr 180
Qy      6136  TGG-----ATTAGAACTGGAAAGCCCCCAAGTGCTCTCTCTTCAGAC 6174
Db      181  TrpGlnGlnArgValGlnLysAspIleArgThrGlySerProLysCysSerProSerAsp 200
Qy      6175  CTAGATTACCTAAATAAGTTTCACCTGAGAAAGGAAACAAGAAATATCCTAAATGG 6234
Db      201  LeuAspTyrLeuLysIleSerPheThrGluLysGlyGlnGluAsnIleLeuLysTrp 220
Qy      6235  GTAATCGTATGCTTCGGGAAATGGTATATTATGAGGCTCGGGTAAACAACACAGGCTCC 6294
Db      221  ValAsnGlyMetSerTrpGlyMetValTyrTyrGlySerGlyLysGlnProGlySer 240
Qy      6295  ATTCTAACTATTCGCCTCAAAATAAAC---CAGCTGAGGCTCCAATGGCTATAGGACCA 6351
Db      241  IleLeuThrIleArgLeuLysIleAsnThrGlnLeuGluProProMetAlaIleGlyPro 260
Qy      6352  AATACGCTCTTGCGGCTCAAGACCCCAACCAACGAGCA----- 6393
Db      261  AsnThrValLeuThrGlyGlnArgProProThrGlnGlyProProHisAsnLeuProVal 280
Qy      6394  -----GGACCATCTCT-----AACATAACTTCTGGA 6420
Db      281  ProGlnGlyProSerProAsnProAspIleThrGlnSerAspTyrAsnIleThrSerGly 300
Qy      6421  TCAGACCCCACT-----GAGTCTAACAGCAGCACTAAATGGGGGCAAAA 6465
Db      301  SerAspProThrAsnThrProArgAsnGluSerAsnSerThrThrLysMetGlyAlaLys 320
Qy      6466  CTTTITAGCTCATCCAGGAGCTTTTCAAGCTCTTAAGCTCCAGCACTCCAGAGGCTACC 6525
Db      321  LeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThr 340
Qy      6526  TCTTCTTTGGCTATGCTTAGCTTCGGGGCCACCTTACTATGAAGGAATGGCTAGAAGA 6585
Db      341  SerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAlaArgArg 360
Qy      6586  GGGAAATTCATGTGACAAAGAACATAGACACCAATGCATGGGGATCCCAAAATPAG 6645
Db      361  GlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGlnAsnLys 380
Qy      6646  CTTACCCCTTACTAGGTTTCTGGAAAAGGCCGTGCATAGGAAGGTTCCCCCATCCAC 6705
Db      381  LeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProProSerHis 400
Qy      6706  CAACACCTTTGTACCACTGAAGCCTTTAATCAAAACCTCTGAGAGTCAATATCTGGTA 6765
Db      401  GlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyrLeuVal 420
Qy      6766  CCTGGTTATGACAGGTGGGCGCATGTAATACCTGATTAACCCCTGTGTTCACCTTG 6825
Db      421  ProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeu 440
Qy      6826  GTTTTAAACCAACTAAGTTTTTGCATTATGTGTCAAATTTCTCCCGCAGTGTATTAC 6885
Db      441  ValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrTrp 460
Qy      6886  TATCCCGAAAAGCAATCCTTTGATGTAATGACTTACAGAAATCATCGACAAAAGAGAA 6945
```

```
Db      461  TyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLysArgGlu 480
Qy      6946  CCCATATCTCTGACACTGCTGCTGTGATGCTCGGAGTTCGGAGTCGACAGGTGTAGGAACA 7005
Db      481  ProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyValGlyThr 500
Qy      7006  GGAAACAGCTGCGCTGCTCAGCGGACACACAGCAGCTAGAAACAGGACTTAGTAACCTACAT 7065
Db      501  GlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHis 520
Qy      7066  CGAATTGTTAACAGAAGATCTCCAAGCCCTAGAAAAATCTGTCAAGTAACCTGGAGGAATCC 7125
Db      521  ArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluSer 540
Qy      7126  CTAACCTCTCTATCTCAAGTAGTCCTACAGAATAGAGAGGGTTAGATTATTATTCTTA 7185
Db      541  LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu 560
Qy      7186  AAAGAAAGGAGGATTATGTGTAGCTTTGAAGAGGAATGCTGTTTTTATGTGGATCATTTCA 7245
Db      561  LysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAspHisSer 580
Qy      7246  GGGGCCATCAGACACTCCATGAACAAGCTTAGAAGAGGTTGGAGAGCGCTCGAAGGGAA 7305
Db      581  GlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArgGlu 600
Qy      7306  AAGAAACTACTCAAGGGTGGTTGAGGATGGTTCAACAGGCTCTCTTTGGTTGGCTACC 7365
Db      601  LysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeuAlaThr 620
Qy      7366  CTACTTTCTGCTTTAAACAGGACCTTTAATAGTCTCTCTCTCTCTCAGAGTTGGGCCCA 7425
Db      621  LeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrValGlyPro 640
Qy      7426  TGTATTATTAAACAAGTTAATTCCTTCATTAGAGAACGAATAAGTGCAGTCCAGATCATG 7485
Db      641  CysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMet 660
Qy      7486  GTACTTAGACAACAGTACCAAGCCGCTCTAGCAGGAGGAGCTGGCCGTAGCTCTAC 7542
Db      661  ValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg---LeuTyr 678
```

## RESULT 10

```
US-10-029-656-2
; Sequence 2, Application US/10029656
; Publication No. US20030175904A1
; GENERAL INFORMATION:
; APPLICANT: Patience, Clive
; APPLICANT: Oldmixon, Beth
; APPLICANT: Ericsson, Thomas
; TITLE OF INVENTION: Molecular Sequence of Pig Endogenous Retrovirus Receptor and Meth
; FILE REFERENCE: 329579-3
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/10/029,656
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Viral
; US-10-029-656-2
```

```
Alignment Scores:
Pred. No.: 4.64e-201 Length: 653
Score: 3147.50 Matches: 593
Percent Similarity: 93.6% Conservative: 18
Best Local Similarity: 90.8% Mismatches: 27
Query Match: 21.5% Indels: 15
DB: 4 Gaps: 3
```

US-10-723-552-3 (1-8132) x US-10-029-656-2 (1-653)

QY 5620 ATGCATCCACGTTAAACCGCGGCCACTCCCGATTCGGGGTGAAGCGCAAGAAAGACTG 5679  
| | | | | : : : : :  
Db 1 MethisProThrLeuSerArgArgHisLeuProIleArgGlyLysProLysArgLeu 20  
QY 5680 AAAATCCCTTAAGCTTCGCTCCATCGCTGGTTCCTTACTCTGTCAATAAAGCTCTCAG 5739  
| | | | | : : : : :  
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40  
QY 5740 ACTAATGGTATCGCATAGGACAGCAGCTGAACCTCCCATAAACCTTATCTCCACCTGG 5799  
| | | | | : : : : :  
Db 41 ValAsnGlyLysArgLeuValAsnSerProAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 5800 TTAATTAATCACTCCGGCACAGGTATTAATATCAACAACAACCTCAAGGGGAGGCTCTCTTA 5859  
| | | | | : : : : :  
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80  
QY 5860 GGAACCTGGTGGCTGATCTATAGCTTGGCTCAGATCAGTATTCCTAGTCTG----- 5913  
| | | | | : : : : :  
Db 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100  
QY 5914 ---ACCTCAACCCAGATATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACCAACA 5970  
| | | | | : : : : :  
Db 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120  
QY 5971 AATAATGGAAACAATTCGGGAATCCAGAGATTTCTTTTGTAAACAATGGAATCTGTGA 6030  
| | | | | : : : : :  
Db 121 AsnAsnGluLutyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140  
QY 6031 ACCTTAATGATGATATTTGGAATGGCAACCTCTCAGCAGGATAGGGTAAGTTTCT 6090  
| | | | | : : : : :  
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160  
QY 6091 TATGTCACAACCTATACCACTCTGGACATTTAATTAC-----CTGACC 6135  
| | | | | : : : : :  
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180  
QY 6136 TGG-----ATTGAACCTGGAAGCCCAAGCTCTCTCTCCAGAC 6174  
| | | | | : : : : :  
Db 181 TrpGlnGluArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200  
QY 6175 CTAGATTACTTAAATAAGTTTCACTGAGAAAGGAAACAAGAAATATCTTAAATGG 6234  
| | | | | : : : : :  
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleLeuLysTrp 220  
QY 6235 GTAATAATGATGCTTTGGGAATGGTATATTATGGAGGCTCGGTAAACAACAGAGCTCC 6294  
| | | | | : : : : :  
Db 221 ValHisGlyMetSerTrpGlyMetValTyrTyrGlySerGlyLysGlnProGlySer 240  
QY 6295 ATTCTAACTATTCCGCTCAAAATAAACACAGCTCGAGCTCCCAATGGCTATAGGACCAAT 6354  
| | | | | : : : : :  
Db 241 IleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMetAlaIleGlyProAsn 260  
QY 6355 ACGGCTTTCAGGGTCAAGACCCCAACCCAGGACCGAGGACCATCTCTTAACATACT 6414  
| | | | | : : : : :  
Db 261 ThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyThrSerSerAsnIleThr 280  
QY 6415 TCTGGATCAGACCCCACTAGTCTAACACGACACTAAATGGGGCAAACTTTTACG 6474  
| | | | | : : : : :  
Db 281 SerGlySerAspProThrGluSerAsnSerThrThrLysMetGlyAlaLysLeuPheSer 300  
QY 6475 CTATCCAGGGAGCTTTTCAAGCTCTTAACCTCACGACTCCAGAGGCTACCTCTCTCTGT 6534  
| | | | | : : : : :  
Db 301 LeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThrSerSerCys 320  
QY 6535 TGGCTATGCTTACGCTTCGGGCCCACTTATATGAAGGAATGGCTAGAGAGGAAATTC 6594  
| | | | | : : : : :  
Db 321 TrpLeuCysLeuAlaLeuGlyProProTyrTyrGlyMetAlaArgGlyLysPhe 340  
QY 6595 AATGTCAAAAAGAACATAGACCAATGCACATGGGGATCCCAAAATAGCTTACCTT 6654  
| | | | | : : : : :  
Db 341 AsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGlnAsnLysLeuThrLeu 360

QY 6655 ACTGAGGTTTCTGAAAAAGGCACCTGCATAGGAAAGGTTCCCCATCCCAACACACTT 6714  
| | | | | : : : : :  
Db 361 ThrGluValSerGlyLysGlyThrCysIleGlyLysValProProSerHisGlnHisLeu 380  
QY 6715 TGTAAACCACTGAAGCCTTTAATCAAACTCTGAGAGTCAATATCTGGTACCTGGTTAT 6774  
| | | | | : : : : :  
Db 381 CysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyrLeuValProGlyTyr 400  
QY 6775 GACAGTGGTGGCATGTAATACTGGAATTAACCCCTTGTCTTCCACTTGGTTTAAAC 6834  
| | | | | : : : : :  
Db 401 AspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeuValPheAsn 420  
QY 6835 CAAACTAAAGATTTTGCATTTATGTCCTCAAAATGTTCCCGAGTGTATTAATCCCGAA 6894  
| | | | | : : : : :  
Db 421 GlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrTyrProGlu 440  
QY 6895 AAGCAATCTTCATCAATATGACTACAGAAATCATCGAAGAGAGAGAACCCATATCT 6954  
| | | | | : : : : :  
Db 441 LysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLysArgGluProIleSer 460  
QY 6955 CTGACACTTGTCTGATGCTCGGACTTGGAGTGGCAGCAGGTGTAGGAAACAGACAGCT 7014  
| | | | | : : : : :  
Db 461 LeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIleGlyValGlyThrGlyThrAla 480  
QY 7015 GCCCTGCTCACGGGACCCAGCAGCTAGAAACAGGACTTAGTAACTTACATCGAAATTGA 7074  
| | | | | : : : : :  
Db 481 AlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHisArgIleVal 500  
QY 7075 ACAGAAGATCTCAAGCCCTAGAAAAATCTGTCTAGTAACTCTGGAGGAATCCCTAACTCC 7134  
| | | | | : : : : :  
Db 501 ThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluGluSerLeuThrSer 520  
QY 7135 TTATCTGAAGTAGTCTCAGAAATAGAGAGGTTAGATTATTATTCTTAAAGAGGA 7194  
| | | | | : : : : :  
Db 521 LeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeuPheLeuLysGluGly 540  
QY 7195 GGATATTGTAGTACCTTGAAGGAGGAATGCTGTTTTTTTGTGATCATTCAGGGGCCATC 7254  
| | | | | : : : : :  
Db 541 GlyLeuCysValAlaLeuLysGluGluCysPheTyrValAspHisSerGlyAlaIle 560  
QY 7255 AGAGACTCATGAACAGCTTAGAGAAAGCTTGGAGAGCGTGAAGGGGAAAGGAAACT 7314  
| | | | | : : : : :  
Db 561 ArgAspSerMetSerLysLeuArgGluArgLeuGluLysArgArgGluLysGluThr 580  
QY 7315 ACTCAAGGCTGTTGAGGGATGTTCAACAGCTCTTTTGGTGGCTACCTACTTCT 7374  
| | | | | : : : : :  
Db 581 ThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeuAlaThrLeuLeuSer 600  
QY 7375 GCTTTAAACAGGACCCCTTAATAGTCTCTCTCTTACTCACAGTTGGGCCATGTATTAT 7434  
| | | | | : : : : :  
Db 601 AlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrValGlyProCysIleIle 620  
QY 7435 AACAGTTAATTCCTTCATTATAGAGAACGAATAAGTGCAGTCCAGATCATGTGTTAGTA 7494  
| | | | | : : : : :  
Db 621 AsnLysLeuIleAlaPheIleArgGluArgGlyIleSerAlaValGlnIleMetValLeuArg 640  
QY 7495 CAACAGTACCAAGCCGCTAGCAGGGAAGCTGGCCGC 7533  
| | | | | : : : : :  
Db 641 GlnGlnTyrGlnSerProSerArgGluAlaGlyArg 653

## RESULT 11

US-10-029-656-4

; Sequence 4, Application US/10029656

; Publication No. US20030175904A1

; GENERAL INFORMATION:

; APPLICANT: Patience, Clive

; APPLICANT: Oldmixon, Beth

; APPLICANT: Erickson, Thomas

; TITLE OF INVENTION: Molecular Sequence of Pig Endogenous Retrovirus Receptor and Met

; TITLE OF INVENTION: Use

; FILE REFERENCE: 329579-3

; CURRENT APPLICATION NUMBER: US/10/029,656

```
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/60/285,103
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Viral
US-10-029-656-4

Alignment Scores:
Pred. No.:      2,94e-200      Length:      653
Score:          3135.50      Matches:      591
Percent Similarity: 93.1%      Conservative: 17
Best Local Similarity: 90.5%      Mismatches:  30
Query Match:     21.4%      Indels:      15
DB:              4          Gaps:          3

US-10-723-552-3 (1-8132) x US-10-029-656-4 (1-653)
Qy      5620 ATGCATCCCAAGTAAACCGCGCCACCTCCCGATTCCGGGTGGAAAGCCGAAAGACTG 5679
Db      1 MetHisProThrLeuSerArgArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20
Qy      5680 AAAATCCCTTAAGCTCGCTCGATCGCGTGTCTTACTCTGTCAATAAATCTCAG 5739
Db      21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
Qy      5740 ACTAATGGTATGCGATAGAGACAGCTCAACTCCCAATAAACCTTTATCTCTCACCTGG 5799
Db      41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
Qy      5800 TTAATTAAGTACCTCCGCGCACAGGTATTAATATCAACAACACTCAAGGGGGAGGTCTCTTTA 5859
Db      61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
Qy      5860 GGAACCTGGTGGCTGATCATAGCTTTGGCTCGATCAGTATTCCTAGTCTG----- 5913
Db      81 GlyThrTrpProGluLeuTyValCysLeuArgSerValIleProGlyLeuAsnAsp 100
Qy      5914 ---ACCTCACCCAGATATCTCTCATGCTCACGGATTTATGTTGCCCGCAGGACACCA 5970
Db      101 GlnAlaThrProProAspValLeuArgAlaTyGlyPheTyValCysProGlyProPro 120
Qy      5971 AATAATGGAACAATCCGGAATCCAGAGATTTCTTTTGTAAACAATGGAATGTA 6030
Db      121 AsnAsnGluGluTyCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
Qy      6031 ACCTTAATGATGATATTGGAATGGCCACCTCTCAGCAGGATAGGTAAGTTTCT 6090
Db      141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnAspArgValSerTySer 160
Qy      6091 TATGTCAACACTATACAGCTCTGGACAATTAATTAC-----CTGACC 6135
Db      161 PheValAsnAsnProThrSerTyArgGlnPheAsnTyArgHisGlyArgTrpLysAsp 180
Qy      6136 TGG-----ATTAGAACTGGAAGCCCAAGTCTCTCTCTTCAGAC 6174
Db      181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
Qy      6175 CTAGATTACCTAAATAAGTTTCACGTAGAAAGGAAACAAAGAAATATCCTAAATGG 6234
Db      201 LeuAspTyLeuLysIleSerPheThrGluLysGlyLysGlnGlnIleLeuLysTrp 220
Qy      6235 GTAATCGTATGCTTCGGGAAATGGTATATTATCGAGGCTCGGTAAACAACACGAGCTCC 6294
Db      221 ValAsnGlyMetSerTrpMetValTyTyGlySerGlyLysGlnProGlySer 240
Qy      6295 ATTCTAACTATTCCCTCAAAATAAACAGCTGGAGGCTCCAATGGCTATAGGACCAAT 6354
Db      241 IleLeuThrIleArgLeuLysIleAsnGlnLeuGluProMetAlaIleGlyProAsn 260
```

```
Qy      6355 ACGGTCTTGACGGGTCAAGAGACCCCAACCCCAAGGACCCAGGACCATCTCTTAACATAACT 6414
Db      261 ThrValIleuThrGlyGlnArgProProThrGlnGlyProGlyProSerSerAsnIleThr 280
Qy      6415 TCTGGATCAGACCCCACTGAGTCTAAACAGCACCACTAAAAATGGGGGCAAACTTTTTCAGC 6474
Db      281 SerGlySerAspProThrGluSerAsnSerThrThrLysMetGlyAlaLysLeuPheSer 300
Qy      6475 CTATCCAGGAGCTTTTCAAGCTCTTAACCTCACGACTCCAGAGGCTACCTCTCTTTGT 6534
Db      301 LeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThrSerSerCys 320
Qy      6535 TGGCTATGCTTAGCTTCGGGCCCCACCTTACTATGAAGGAATGGCTAGAACAGGGAATTC 6594
Db      321 TrpLeuCysLeuAlaLeuGlyProProTyTyGlyMetAlaArgArgGlyLysPhe 340
Qy      6595 AATGTGACAAAAAGAACATAGAGACCAATGCACATGGGGATCCCAAAATAAGCTTTACCCCT 6654
Db      341 AsnValThrLysGlyHisArgAspProCysThrTrpGlySerGlnAsnLysLeuThrLeu 360
Qy      6655 ACTGAGTCTTCTGAAAGAGCCTCGCATAGGAAAGTTCCCCCATCCCAACCAACACTT 6714
Db      361 ThrGluValPheGlyLysGlyThrCysIleGlyLysValProProSerHisGlnHisLeu 380
Qy      6715 TGTAAACACACAGCAAGCCTTAACTCAACCTCTGAGAGTCAATATCTGGTACCTGGTTAT 6774
Db      381 CysAsnHisThrGluAlaPheAsnArgThrSerGluSerGlnTyLeuValProGlyTy 400
Qy      6775 GACAGTGGTGGGCATGTAATCTGGAATTAACCCCTTGTTGTTCCACCTTGGTTTTTAAC 6834
Db      401 AspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeuValPheAsn 420
Qy      6835 CAACTAAAGATTTTTCATATTATGTCCTCAAAATTTGTTCCCGAGTGTTATTAATTCGAA 6894
Db      421 GlnThrLysAspPheCysIleMetValGlnIleValProArgValTyTyTyProGlu 440
Qy      6895 AARGCAATCTTCATCAATATGACTACAGAAATCATCGAACAAAGAGAGAACCCATATCT 6954
Db      441 LysAlaIleLeuAspGluTyAspTyArgAsnHisArgGlnLysArgGluProIleSer 460
Qy      6955 CTGACACTTGTCTGTGCTCGGACTTTGGAGTGGCAGCAGGTCTAGGAAACAGGAACAGCT 7014
Db      461 LeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyValGlyThrAla 480
Qy      7015 GCCTGTGTCAGGACACAGCAGCTAGAAACAGGACTTAGTAACCTACATCGAATTGTA 7074
Db      481 AlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHisArgIleVal 500
Qy      7075 ACAGAAAGATCTCCAAGCCCTAGAAAAATCTGCAGTAACCTGGAGGAATCCCTAACCTCC 7134
Db      501 ThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluSerLeuThrSer 520
Qy      7135 TTATCTGAAGTACTCTACAGAAATAGAAGGGGTAGATTATTATTCTTAAAGGAAGA 7194
Db      521 LeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuLysGluGly 540
Qy      7195 GGAATTATGCTAGCTTGAAGGAGGAATGCTGTTTTTATGTGATCATTTAGGGGCCATC 7254
Db      541 GlyLeuCysValAlaLeuLysGluGlyCysCysPheTyValAspHisSerGlyAlaIle 560
Qy      7255 AGAGACTCCCATGAACAGCTTACAGAAAGGTTGAGAGAGCGTCGAGGGAAGGAAAGAACT 7314
Db      561 ArgAspSerMetSerLysLeuArgGluArgLeuLysArgArgGlyLysGluThr 580
Qy      7315 ACTCAAGGGTGGTTGAGGATGTTTCAACAGGTCTCTTTTGGTTGGCTACCTACTTTCT 7374
Db      581 ThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeuAlaThrLeuSer 600
Qy      7375 GCTTTAACAGGACCTTAATAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7434
Db      601 AlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrValGlyProCysIleIle 620
Qy      7435 AACAAAGTTAATTCCTTCATTAGAGAACGAATAAGTGCACTCCAGATCATGGTACTTAGA 7494
```

```
Db 621 AsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMetValLeuArg 640
QY 7495 CAACAGTACCAAGCCCGTCTAGCAGGGAAGCTGGCCGC 7533
Db 641 GlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 653

RESULT 12
US-10-441-949-43
; Sequence 43, Application US/10441949
; Publication No. US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodrigo, Allen
; APPLICANT: Ross, Howard A.
; APPLICANT: Mullins, James I.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; PRIOR FILING DATE: 2003-05-19
; PRIOR FILING DATE: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-43

Alignment Scores:
Pred. No.: 3 65e-183 Length: 661
Score: 2879.50 Matches: 541
Percent Similarity: 88.3% Conservative: 40
Best Local Similarity: 82.2% Mismatches: 56
Query Match: 19.7% Indels: 21
DB: 4 Gaps: 6

US-10-723-552-3 (1-8132) x US-10-441-949-43 (1-661)
QY 5620 ATGCATCCACGTTAAACCGGCCACCTCCCGATTCCGGGTGGAAGCCGAAAGACTG 5679
Db 1 MetHisProThrLeuSerArgHisLeuProIleArgGlyLysProLysArgLeu 20
QY 5680 AAAATCCCTTAAGCTTCGCTCCATCGGTGGTTCCTTACTCTGTCAATAACCTCTCAG 5739
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
QY 5740 ACTAATGGTATGGCATAGGACACCTGAACCTCCCAATAAACCTTATCTCACCCTGG 5799
Db 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
QY 5800 TTAATTAAGTACCTCCGCGCACAGGTATTAATATCAACACACTCAAGGGGAGGCTCCTTTA 5859
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
QY 5860 GGAACTGGTGGCTGATCTATACGTTTGGCTCAGATCAGTTATTCCTAGTCTG----- 5913
Db 81 GlyThrTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY 5914 ---ACCTCAACCCACAGATATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACCA 5970
Db 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
QY 5971 AATAATGGAACCAATTCGGAATCCAGAGATTTCTTTTGTAAACAAATGGAAGTGTGA 6030
Db 121 AsnAsnGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
QY 6031 ACCTCTAATGATGATATTGGAAATGGCCAACTCTCAGCAGATAGGTAAGTTTCT 6090
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160
QY 6091 TATGTCCAACACCTATACACAGCTCTCGACATTTAAATTAC-----CTGACC 6135
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
QY 6136 TGG-----ATTAGAACTGGGAAGCCCAAGTCTCTCTCTTCAGAC 6174
Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysAsnSerLeuAsp 200
QY 6175 CTAGATTACTTAAATAAGTTTCACTGAGAAAGGAACAAAGAAATATCTCTAAATGG 6234
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyGlnGlnAsnIleGlnLysTrp 220
QY 6235 GTAAATGGTATGCTTGGGAATGGTATATATGAGAGGCTCGGTAAACCAACAGGCTCC 6294
Db 221 ValAsnGlyMetSerTrpGlyIleValTyrGlyGlySerGlyArgLysGlySer 240
QY 6295 ATTCTAACTATTGCGCTCAAAATA---AACCAGCTGGAGCTCCTCAATGGCTATAGACCA 6351
Db 241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProValAlaIleGlyPro 260
QY 6352 AATACGGTCTTGACGGGTCAAGACCCCAACCCAA-----GGACCAGGACCA 6399
Db 261 AsnLysGlyLeuAlaGluGlnGlyProIleGlnGlnArgProSerProAsnPro 280
QY 6400 TCCTCT---AACATAACTTCTGGATCAGACCCCACTGAGTCTAACAGCAGCTAAATG 6456
Db 281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300
QY 6457 GGGGCAAACTTTTGTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCAGACTCCA 6516
Db 301 GlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320
QY 6517 GAGGCTACCTCTCTTGTGGCTATGCTTAGTCTCGGGCCCACTTACTATGAGGAATG 6576
Db 321 GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMet 340
QY 6577 GCTAGAAGAGGGAATTCATATGTGACAAAGAACATAGAGACCAATGCATCGGATCC 6636
Db 341 AlaArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySer 360
QY 6637 CAAATAAGCTTACCTTACTGAGGTTCTGGAAGAGGACCTGCATAGAGAAAGTTCCC 6696
Db 361 GlnAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyArgValPro 380
QY 6697 CCATCCCAACACACTTTGTAAACCACTGAAGCTTTAATCAAACTCTGAGAGTCAA 6756
Db 381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400
QY 6757 TATCTGGTACCTGGTTTATGACAGGTGGTGGCATGTAATCTGGATTAAACCTCTGTGT 6816
Db 401 TyrLeuValProGlyTyrAspArgTrpAlaCysAsnThrGlyLeuThrProCysVal 420
QY 6817 TCCACCTTGGTTTAAACCAAACTAAAGATTTTTCATTATGCTCCAAATTTGTCCTCCGA 6876
Db 421 SerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArg 440
QY 6877 GTGTATTACTATCCGAAAGAAACCAATCCTTGTATGATATGACTACAGAAATCATCCACA 6936
Db 441 ValTyrTyrProGluLysAlaValLeuAspGluTyrAspTyrArgTyrAsnArgPro 460
QY 6937 AAGAGAGAACCATATCTCTGACACTTGTGTGATGCTCGGACTTCGAGTGGCAGCAGGT 6996
Db 461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaGly 480
QY 6997 GTAGGAACAGGAACAGCTGCCCTGGTCCAGGGACCAACAGCAGCTAGAAAACAGACTTAGT 7056
Db 481 ValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSer 500
QY 7057 AACCTACATCGAATTTGTAAACAGAAATCTCCAAAGCCCTAGAAAAATCTCTAGTACCTGT 7116
Db 501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520
```

```
QY 7117 GAGGAATCCCTAACCTCTTATCTGAAGTAGTCTCAGAAATACAGAGGGTTAGATTTA 7176
Db 521 GluIuSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 540
QY 7177 TTATTTCTAAAGAGGAGGATTATGTGTAGCTTGAAGGAGGAATCTGTTTTTATGTG 7236
Db 541 LeuPheLeuLysGluGlyGlyLeuCyseValAlaLeuLysGluGlyCysPheTyZVal 560
QY 7237 GATCATTACGGGGCCATCAGAGACTCCATGAAACAGCTTACAGAAAGGTTGGAGACGCT 7296
Db 561 AspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArgArg 580
QY 7297 CGAAGGGAAGGAACCTACTCAAGGCTGTTTGGAGGATGGTTCAACAGCTCTCTTTGG 7356
Db 581 ArgArgGluArgGluAlaAspGlnGlyTirPheGlnGlyTirPheAsnArgSerProTir 600
QY 7357 TTGGCTACCTCTCTTCTGCTTTAAACAGGACCTTAAATAGTCTCTCTCTTACTCACA 7416
Db 601 MetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuThr 620
QY 7417 GTTGGGCCATGTATTATTAACAAGTTAATTGCTTCAATAGAGAACGAATAAGTGCAGTC 7476
Db 621 ValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaVal 640
QY 7477 CAGATCATGTGTACTACACACAGTACCAAGCCGCTTACGAGGGAAGCTGGC 7530
Db 641 GlnIleMetValLeuArgGlnGlnTyrGlnGlyLeuLeuSerGlnGlyGluGly 658
```

## RESULT 13

```
US-10-441-949-16
; Sequence 16, Application US/10441949
; Publication No. US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodrigo, Allen
; APPLICANT: Ross, Howard A.
; APPLICANT: Mullins, James I.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; CURRENT FILING DATE: 2003-05-19
; PRIOR FILING DATE: 2003-05-19
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-16
```

```
Alignment Scores:
Pred. No.: 7,86e-183 Length: 660
Score: 2874.50 Matches: 539
Percent Similarity: 88.8% Conservative: 38
Best Local Similarity: 82.9% Mismatches: 52
Query Match: 19.6% Indels: 21
DB: 4 Gaps: 6
```

US-10-723-552-3 (1-8132) x US-10-441-949-16 (1-660)

```
QY 5620 ATGCATCCCAAGTTAAACCGCGCCACCTCCGATTCCGGGTGGAAGCCGAAAGACTG 5679
Db 1 MethHisProThrLeuSerArgArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20
QY 5680 AAAATCCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCAATAACTCTCAG 5739
Db 21 LysIleProLeuSerPheAlaSerIleAlaTirPheLeuThrLeuSerIleThrProGln 40
```

```
QY 5740 ACTAATGGTATGCGCATAGGAGACAGCTGAACTCCCATAAACCCCTTATCTCTCACCTGG 5799
Db 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTirP 60
QY 5800 TTAATTAAGTCACTCCGCGCACAGGTATTAATATCAACAACACTCAAGGGGAGGCTCTTTTA 5859
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
QY 5860 GGAACCTGGTGGCTGATCTATACGTTTGCCTCAGATCAGTATTCTTCTAGTCTG----- 5913
Db 81 GlyThrTirPirProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY 5914 ---ACCTCACCCACAGATATCTCCATGCTCAGGATTTTATGTTGCCAGAGACACCA 5970
Db 101 GlnAlaThrProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
QY 5971 AATAATGGAAAAATTCGCGAAATCCAGAGATTTCTTTTGTAAACAATGGAATCTGTGTA 6030
Db 121 AsnAsnGluGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTirPserCysVal 140
QY 6031 ACCTCTAATGATCGATATTGGAATTCGCAACCTCTCAGCAGATAGGTAAGTTTCT 6090
Db 141 ThrSerAsnAspGlyAsnTirPstTirProValSerGlnGlnAspArgValSerTirSer 160
QY 6091 TATGTCAACACCTATACAGCTCTGGACAAATTTAATTAC-----CTGACC 6135
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTirPLeuAsp 180
QY 6136 TGG-----ATTAGAACTGGAAAGCCCCCAAGTCTCTCTCTTCAGAC 6174
Db 181 TirPirGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
QY 6175 CTAGATTACCTAAATAAGTTTCACTCAGAAAGGAAACAAAGAAATATCTTAATG 6234
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlnAsnIleGlnLysTirP 220
QY 6235 GTAAATGGTATGCTTTGGGAATGGTATATTATGGAGGCTCGGTTAAACAAACACGAGTCC 6294
Db 221 ValAsnGlyMetSerTirPirIleValTyrGlySerGlyArgLysLysGlySer 240
QY 6295 ATTCTAATTCCTCAAAATA---AACCACTGGAGCTCCAAATGGGTATAGGACCA 6351
Db 241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProValAlaIleGlyPro 260
QY 6352 AATACGCTCTTGACGGGTCAAGACCCCAACCCAA-----GGACCCAGGACCA 6399
Db 261 AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGlnArgProSerProAsnPro 280
QY 6400 TCCTCT---AACATAACTCTGGATCAGACCCCACTGAGTCTAACACACAGCTATAAATG 6456
Db 281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300
QY 6457 GGGGCAAAACCTTTTACCTTCATCCAGGAGCTTTTCAAGCTCTTAATCTCAGACTCCA 6516
Db 301 GlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320
QY 6517 GAGCTACCTCTCTTTGTTGGCTATGCTTACCTCGGGCCCACTTACTTATGAGGAATG 6576
Db 321 GluAlaThrSerSerCysTirPLeuCysLeuAlaSerGlyProProTyrTirGluGlyMet 340
QY 6577 GCTAGAAGAGGGAATTCATGTGACAAAGAACATAGACCAATGACATGCGATGGGATCC 6636
Db 341 AlaArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTirPirLysSer 360
QY 6637 CAAAATAAGCTTACCTTACTGAGGTTTCTGGAAGAGGCACTGCATAGGAAAGTTCCC 6696
Db 361 GlnAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyArgValPro 380
QY 6697 CATCCCAACCAACCTTTGTACCACTGAGACCTTTAATCAACCTCTGAGAGTCAA 6756
Db 381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400
QY 6757 TATCTGTACTGTTTATGACAGTGTGGGCATGTAATACTGGATTAAACCCCTTGTGTT 6816
```

Db	401	TyrLeuValProGlyTyraSPargtirPralaCysaenThrGlyLeuThrProCysVal	420
Qy	6817	TCCACCTTGTTTTTAACCAACATAAGATTTTTGCAATTATGTGCCAAATTTGTTCCCGGA	6876
Db	421	SerThrLeuValPheAenGlnThrLysaAspPheCysValMetValGlnIleValProArg	440
Qy	6877	GTGTATTACTATCCGAAAAAGCAATCCTTCATGAATATGACTACAGAAATCATCGACAA	6936
Db	441	ValTyrtYrTyPrOGluLyAlaValLeuAspGluTyraSPtyrArAgtYraAsnArgPro	460
Qy	6937	AAGAGAAACCCATATCTCTCACACTTGCTGTGATGTCGCGACTTGCGAGTGGCGACAGAGT	6996
Db	461	LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGly	480
Qy	6997	GTAGAAACAGAAACAGCTGCCTGGTCAACGGACACACGACGTAGAAACAGGACTTAGT	7056
Db	481	ValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnLeuGluLyysGlyLeuSer	500
Qy	7057	AACCTACATCGAATTTGAACAGAGAGTCTCCAAGCCCTAGAAATCTGTCAGTAACTCG	7116
Db	501	AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLyysSerValSerAsnLeu	520
Qy	7117	GAGGAATCCCTAAACCTCTTATCTCAAGTAGTCTCTACAGATACAGAGGCTTAGATTTA	7176
Db	521	GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu	540
Qy	7177	TTATTTCTAAAGAAGGAGGATTATGTGTAGCTTGAAGGAGGAATGCTGTTTTTATGTG	7236
Db	541	LeuPheLeuLyysGluGlyGlyLeuCysValAlaLeuLyysGluGluCysCysPheTyVal	560
Qy	7237	GATCATTTCAGGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGGTTGGAGAAGCGT	7296
Db	561	AspHisSerGlyAlaIleArgaSPSerMetSerLyseuArgGluArgLeuGluArgArg	580
Qy	7297	CGAAGGAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGGTTCAACAGGCTCTCTTTGG	7356
Db	581	ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTip	600
Qy	7357	TTGGCTACCTTACTTCTGCTTTAACAGGACCCCTTAATAGTCTCTCTCTGTTACTCACA	7416
Db	601	MetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThr	620
Qy	7417	GTTGGGCCATGTATTATTAAACAAGTAATTGCCTTCATTAGAGAACCAATAGTGCACTG	7476
Db	621	ValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaVal	640
Qy	7477	CAGATCATGTACTTAGACAAACAGTACAA	7506
Db	641	GlnIleMetValLeuArgGlnGlnTrGln	650

## RESULT 14

```

US-10-441-949-8
; Sequence 8, Application US/10441949
; Publication No. US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodrigo, Allen
; APPLICANT: Ross, Howard A.
; APPLICANT: Mullins, James I.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

```

;
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
; US-10-441-949-8

Alignment Scores:
Pred. No.: 1.25e-182 Length: 660
Score: 2871.50 Matches: 539
Percent Similarity: 88.6% Conservative: 37
Best Local Similarity: 82.9% Mismatches: 53
Query Match: 19.6% Indels: 21
DB: 4 Gaps: 6

US-10-723-552-3 (1-8132) x US-10-441-949-8 (1-660)

Qy 5620 ATGCATCCCAAGTTAAACCGCGCCACCTCCCGATTTCGGGTGGAAGCGGAAAGACTG 5679
Db 1 MethiProThrLeuSerArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20
Qy 5680 AAAATCCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCAATAAACCCTCTCAG 5739
Db 21 LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
Qy 5740 ACTAATGGTATGCGCATAGGAGACAGCCTGAACTCCATAAACCCCTATCTCTCACCTGG 5799
Db 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
Qy 5800 TTAATTACTGACTCCGCGCACAGGTATTAATATCAACAACACTCAAGGGAGGCTCCTTTA 5859
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
Qy 5860 GSAACCTGGTGGCTCATCTATACGTTTCCTCCAGATCAGTTATTCCTAGTCTG----- 5913
Db 81 GlyThrTrpProGluLeuTyValCysLeuArgSerValileProGlyLeuAsnAsp 100
Qy 5914 ---ACCTCACCCCGATATCTCCATGCTCACGATTTTATGTTGGCCAGCACACCA 5970
Db 101 GlnAlaThrProProAspValLeuArgAlaTyGlyPheTyValCysProGlyProPro 120
Qy 5971 AATAATGGAAACATTCGGAAATCCAGAGATTTCTTTGTGTAAACAATGGAACTGTGA 6030
Db 121 AsnAsnGluGluTyCysGlyAsnProGlnAspPhePheCysGlnTrpSerCysVal 140
Qy 6031 ACCTCTAATGATGATATTGGAAATGCCAACCTCTCAGCAGATAGGGTAAGTTTCTT 6090
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160
Qy 6091 TATGTCACACCTATACCAAGCTCTCGACAAATTAAATAC-----CTGACC 6135
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
Qy 6136 TGG-----ATTAGAACTGGAAAGCCCAAGTGCTCTCTTCAGAC 6174
Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
Qy 6175 CTAGATTACCTAAAAATAAGTTTCACTGAGAAAGGAAAAACAAGAAATAATCTTAAATGG 6234
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrp 220
Qy 6235 GTAATGGTATGCTTTGGGAATGGTATATATGAGAGGCTCGGGTAAACAACACAGCTCC 6294
Db 221 ValAsnGlyMetSerTrpGlyIleValTyTyTyGlySerGlyArgLysLysGlySer 240
Qy 6295 ATTCTAACTATTGCCTCAAAATA---AACCAAGCTGGAGCTCCCAATGGCTATAGGACCA 6351
Db 241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProValAlaIleGlyPro 260
Qy 6352 AATACGGTCTTCAGGGTCAAAAGACCCCAACCCAA-----GGACCAAGACCA 6399
Db 261 AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGluGlnArgProSerProAsnPro 280
Qy 6400 TCCTCT---AACATAACTTCTGGATCAGACCCCACTGAGTCTAACAGCAGCACTAAATG 6456
Db 281 SerAspTyAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300

```





Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180  
 QY 6136 TGG-----ATTAGAACTGGAAGCCCAAGTGTCTCTCCCTTCAGAC 6174  
 Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200  
 QY 6175 CTAGATTACCTAAAATAAGTTTCACCTGAGAGAGAAACAGAAAATATCTCTAAATGG 6234  
 Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrp 220  
 QY 6235 GTAATCGTATGCTTCGGGAATGCTATATTATGAGGCTCGGTAAACAACACAGGCTCC 6294  
 Db 221 ValAsnGlyIleSerTrpGlyIleValTyrTrpGlySerGlyArgLysLysGlySer 240  
 QY 6295 ATTCTAACTATTGCGCTCAAAATA---AACACAGCTGGAGCTCCCAATGGCTATAGGACCA 6351  
 Db 241 ValLeuThrIleArgLeuArgGlyLeuThrGlnMetGluProProValAlaIleGlyPro 260  
 QY 6352 AATACGGTCTTGACGGGTCAAGACCCCAACCCCA-----GGACACGAGACCA 6399  
 Db 261 AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGluGlnArgProSerProAsnPro 280  
 QY 6400 TCCTCT---AACATACTCTGATCAGACCCCACTGAGTCTAACACACGACTAAATG 6456  
 Db 281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300  
 QY 6457 GGGGCAAAACTTTTAGCCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCACGACTCCA 6516  
 Db 301 GlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320  
 QY 6517 GAGGCTACCTCTCTTGTGTTGGCTATGCTTAGCTTCGGGCCCACTTACTATGAAGGAATG 6576  
 Db 321 GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTrpGluGlyMet 340  
 QY 6577 GCTAGAGAGGGAATCAATGTGACAAAGACATAGACCAATGCACATCGGGATCC 6636  
 Db 341 AlaArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySer 360  
 QY 6637 CAAATAAGCTTACCTTACTGAGGTTCGTGAAAAGGACCTGCATAGGAAAGTTCCC 6696  
 Db 361 GlnAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyMetValPro 380  
 QY 6697 CCAATCCACCAACACCTTTGTAAACCACTGAAAGCCTTTAATCAAACTCTGAGAGTCAA 6756  
 Db 381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400  
 QY 6757 TATCTGGTACCTGGTTATGACAGCTGGTGGGCACTGTAATCTGATTAACCCCTTGTGT 6816  
 Db 401 TyrLeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVal 420  
 QY 6817 TCCACCTTGGTGTTTTAAACCAACTAAAGATTTTTCATTATGTCCTCAAAATGTTCCCGCA 6876  
 Db 421 SerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArg 440  
 QY 6877 GTGTATTACTATCCCGAAAAGCAATCCTTTGATGAATATGACTACAGAAATCATCGACAA 6936  
 Db 441 ValTyrTyrProGluLysAlaValLeuAspGluTyrAspTyrArgTyrAsnArgPro 460  
 QY 6937 AAGAGAACCCCATATCTCTGACACTGCTGTGATGCTGGACTGGAGTGGCAGCAGGT 6996  
 Db 461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaGly 480  
 QY 6997 GTAGGAACAGGAACAGCTGCGCTGCTCAGGGACACAGCAGCTAGAAAACAGGACTTAGT 7056  
 Db 481 ValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSer 500  
 QY 7057 AACCTACATCGAATTCTAACAGAAGATCTCCAAGCCCTAGAAAATCTGTCAATAACCTG 7116  
 Db 501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520  
 QY 7117 GAGGAATCCCTAACCTCTTATCTGAAGTAGTCTTACAGATAGAGAGGGTTAGATTTA 7176  
 Db 521 GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgGlyLeuAspLeu 540

QY 7177 TTATTTCTAAAGAGGAGGATTATGTAGCCTTGAAGAGGAATGCTGTTTTATGTG 7236  
 Db 541 LeuPheLeuLysGluGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrVal 560  
 QY 7237 GATCATTCAGGGGCCATCAGAGACTCATCAACAAAGCTTAGAGAAAAGTTGGAGAACGT 7296  
 Db 561 AspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArgArg 580  
 QY 7297 CGAAGGMAAGCAAACTACTCAAGGGTGGTTTGAGGGATGGTTCACAGGTCTCTTTGG 7356  
 Db 581 ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrp 600  
 QY 7357 TTGGCTACCTCTTCTTCTTTAAACAGGACCTTAAATAGTCTCTCTCTCTCTCTCACA 7416  
 Db 601 MetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuThr 620  
 QY 7417 GTTGGGCCATGTATTATTAACTTAATTCCTTCAATTAGAGAACGAATAGTCAGTC 7476  
 Db 621 ValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaVal 640  
 QY 7477 CAGATCATGTACTTATAGACAAACAGTACCAA 7506  
 Db 641 GlnIleMetValLeuArgGlnGlnTyrGln 650

Search completed: February 14, 2006, 17:17:01

Job time : 1161.38 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 16:19:01 ; Search time 7.13096 Seconds  
(without alignments)  
2992.933 Million cell updates/sec

Title: US-10-723-552-3  
Perfect score: 14636  
Sequence: 1 GCGTGTGTACGACTGTGGG.....CTGTTTGCATCAAAAAAAA 8132

Scoring table: BLOSUM62  
-O= /abs/ABSWEB.spool/US10723552/runat.14022006.125151.13368/app.query.fasta\_1  
-DB=Published Applications\_AA\_New -QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-MAXLEN=200000000 -HOST=abes02p  
-USER=US10723552.@CGN.1.1.17.0runat.14022006.125151.13368 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Searched: 97014 seqs, 1312538 residues  
Total number of hits satisfying chosen parameters: 194028

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSWEB.spool/US10723552/runat.14022006.125151.13368/app.query.fasta\_1  
-DB=Published Applications\_AA\_New -QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-MAXLEN=200000000 -HOST=abes02p  
-USER=US10723552.@CGN.1.1.17.0runat.14022006.125151.13368 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA\_New:  
1: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pcp:\*  
2: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pcp:\*  
3: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pcp:\*  
4: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pcp:\*  
5: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB.pcp:\*  
6: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pcp:\*  
7: /cgn2\_6/prodata/1/pubpaa/US11\_NEW\_PUB.pcp:\*  
8: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1735	11.9	1189	US-10-821-234-1209	Sequence 1209, Ap
2	1646.5	11.2	1150	US-10-821-234-1083	Sequence 1083, Ap
3	1358	9.3	537	US-11-129-442-47	Sequence 47, Appl
4	1267.5	8.7	654	US-10-510-947-6	Sequence 6, Appl
5	431	2.9	912	US-11-042-988-12	Sequence 12, Appl
6	413.5	2.8	540	US-10-821-234-1395	Sequence 1395, Ap
7	301	2.1	562	US-10-507-928-10	Sequence 10, Appl
8	301	2.1	562	US-10-507-928-12	Sequence 12, Appl
9	301	2.1	562	US-11-029-465-10	Sequence 10, Appl

10	287.5	2.0	561	7	US-11-029-465-12	Sequence 12, Appl
11	250	1.7	229	7	US-11-014-629-1	Sequence 1, Appl
12	217.5	1.5	210	7	US-11-022-562-221	Sequence 221, App
13	210.5	1.4	422	7	US-11-230-251-26	Sequence 26, Appl
14	199.5	1.4	2760	7	US-11-124-367A-444	Sequence 444, App
15	199.5	1.4	2803	7	US-11-124-367A-442	Sequence 442, App
16	199.5	1.4	2803	7	US-11-124-367A-445	Sequence 445, App
17	199.5	1.4	2984	7	US-11-124-367A-443	Sequence 443, App
18	199.5	1.4	3027	7	US-11-124-367A-441	Sequence 441, App
19	181.5	1.2	2108	6	US-10-995-561-989	Sequence 989, App
20	181	1.2	1823	6	US-10-995-561-988	Sequence 988, App
21	181	1.2	2102	6	US-10-995-561-990	Sequence 990, App
22	181	1.2	2157	6	US-10-995-561-991	Sequence 991, App
23	174.5	1.2	188	7	US-11-234-786-592	Sequence 592, App
24	167	1.1	14130	7	US-11-175-689-9	Sequence 9, Appl
25	163.5	1.1	416	7	US-11-226-657-136	Sequence 136, App
26	158.5	1.1	16990	7	US-11-175-689-7	Sequence 7, Appl
27	157.5	1.1	1558	6	US-10-329-258-14	Sequence 14, Appl
28	155.5	1.1	2766	6	US-10-877-346-62	Sequence 62, Appl
29	152	1.0	1730	7	US-11-182-016-19	Sequence 19, Appl
30	151.5	1.0	1516	6	US-10-220-824-8	Sequence 8, Appl
31	151.5	1.0	1560	7	US-11-059-982-1	Sequence 1, Appl
32	150.5	1.0	3073	7	US-11-143-980-50	Sequence 50, Appl
33	150	1.0	1717	7	US-11-182-016-20	Sequence 20, Appl
34	148	1.0	1717	7	US-11-182-016-20	Sequence 20, Appl
35	147.5	1.0	1388	6	US-10-821-234-1143	Sequence 1143, Ap
36	147	1.0	1618	6	US-10-984-645-2	Sequence 2, Appl
37	147	1.0	4419	6	US-10-821-234-1155	Sequence 1155, Ap
38	144.5	1.0	400	7	US-11-186-284-155	Sequence 155, App
39	144	1.0	2811	6	US-10-877-346-27	Sequence 27, Appl
40	143.5	1.0	1076	6	US-10-467-657-7916	Sequence 7916, Ap
41	143	1.0	1730	7	US-11-182-016-19	Sequence 19, Appl
42	142.5	1.0	1410	6	US-10-821-234-1050	Sequence 1050, Ap
43	142.5	1.0	1814	6	US-10-877-346-25	Sequence 25, Appl
44	142	1.0	1742	7	US-11-182-016-23	Sequence 23, Appl
45	142	1.0	2011	7	US-11-080-991-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1  
US-10-821-234-1209  
; Sequence 1209, Application US/10821234  
; Publication No. US20050255114A1  
; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Andarmani, Susan  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia  
; FILE REFERENCE: 821A  
; CURRENT APPLICATION NUMBER: US/10/821,234  
; CURRENT FILING DATE: 2004-04-07  
; PRIOR APPLICATION NUMBER: US 60/462,047  
; PRIOR FILING DATE: 2003-04-07  
; NUMBER OF SEQ ID NOS: 1704  
; SOFTWARE: pt\_seq\_genes Version 1.0  
; SEQ ID NO 1209  
; LENGTH: 1189  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(1189)  
; OTHER INFORMATION: Xaa = any amino acid or nothing  
US-10-821-234-1209

Alignment Scores:  
Pred. No.: 1.17e-113 Length: 1189  
Score: 1735.00 Matches: 453  
Percent Similarity: 50.9% Conservative: 192  
Best Local Similarity: 35.8% Mismatches: 462  
Query Match: 11.9% Indels: 160



Db 655 ProLeuThrLys\*\*\*LeuLysProGlySerSerSerGlnLysAlaGluLeuLeuAlaLeu 674  
QY 4212 AGCAAGCTTTGGCGCTGGCGAAGGAAATCCATAAACAATTTATACGACAGCAGGTAT 4271  
Db 675 ThrAlaLeuAenLeuSerLysGlyValAsnIleThrAspSerLysTyr 694  
QY 4272 GCCTTTCGAGTGCACAGTACATGGGGCCATCTATAACAAGGGGGTTCCTTACCTCA 4331  
Db 695 AlaTyrHisIleProArgSerHisAlaAlaIleThrGlnLysArgGlyLeuLeuThrAla 714  
QY 4332 GCAGGAGGGAATAAAGACAAAGAGAAATCTAAGCCATTAGAACCGGTACATTTA 4391  
Db 715 LysGlyThrProIleThrAsnGlyHisLeuIleTyr\*\*\*LeuLeuGlnAlaThrHisLeu 734  
QY 4392 CCAAAAAGGCTAGCTATTATACACTGCTCGACATCAGAAAGCTAAGATCTCATATCC 4451  
Db 735 ProAlaLysAlaGlyValIleHisCys\*\*GlyHisArgThrGlySerAspGluIleSer 754  
QY 4452 AGAGAAACCATGAGTGGTCCAGGGTGGCCAGGAGCAGCCAGGGTGTAACTTCTG 4511  
Db 755 LysGlyAsnArgLysThrAspGluAlaAlaLysGlnAspSerLeuSerProLeuProAla 774  
QY 4512 CCTATATAGAAATGCCAAAGCCCAAGACCCAGACGACAGTACACCTTAGAAGACTGG 4571  
Db 775 ProIleLeuLeuValThrProAlaValProArgTyrProProThr----- 790  
QY 4572 CAAGAGATAAAAAGATAGACCACTCTCTGAGACTCCGGAAGGAGCTGCTATACCTCA 4631  
Db 791 ---GluLysSerSerLeuLeuGlnGlnGlyAlaSerLeuGlnGlyAspTrpIleLys 809  
QY 4632 GATGGGAGAAATCTGCCCCCAAGAAGGGTTAGAAATATGCCAACAGATACATCGT 4691  
Db 810 AsnGlnLysProValLeuProGlnGlnSerLysGluIleLeuThrProLeuHisGln 829  
QY 4692 CTAACCCACTAGCACTAAACACCTGCAGCAGTGTGTCAGA-----ACATCCCT 4742  
Db 830 ProPheHisIleSerAlaCysProLeuTyrLeuLeuLeuLeuProCysPheSerSerPro 849  
QY 4743 TATCATGTTCTGAGGTACAGGAGTGGCTGACTCGGTGTCAAAACATTGTGTGCCCTGC 4802  
Db 850 HisLeuPheIleSerLeu-----LysAspIleThrSerAsnCysArgIleCys 865  
QY 4803 CAGCTGTTAATGCTAATCTTCAGAAATGCCT-----CCAGGAAGAGACTAAG 4853  
Db 866 SerValThrSerSerGlnGlyAlaLeuCysProLeuLeuLeuIleLeuThrTyrGlnLeuArg 885  
QY 4854 GGAAGCCACCCAGCGCTCACTGGAGTGGACTCACTGAGGTAAAGCCGCTAATAC 4913  
Db 886 GlyThrLeuProGlyGluHisIleThrGlnValAsnPheThrHisMetProValLys--- 904  
QY 4914 GGAACAAATACCTATTGGTTTGTAGACACCTTTTCAGATGGGTAGAGGCTTATCCT 4973  
Db 905 LysSerLysTyrLeuLeuThrLeuValAspThrPheSerGly\*\*\*ValGluAlaPhePro 924  
QY 4974 ACTAAGAAAGACTTCAACCGTGGTGTGCTTAAATAAATCTGGAAGAAATTTTCAAGA 5033  
Db 925 ThrProSerGlyLysAlaAlaGluValSerGlnIleLeuValThrGluIleLeuProArg 944  
QY 5034 TTTGGATACCTAAGGTAAATAGGTGACAAATGTCAGCTTTTGTGCCAGGTAACT 5093  
Db 945 PheGlyLeuProGlySerIleGlnSerAspAsnSerProSerPheIleSerGlnIleThr 964  
QY 5094 CAGGAGCTGCCAAGATATTGGGATTTGATTGAAACTGCTATGTCATACAGACCCCAA 5153  
Db 965 GlnGlnValSerGlnSerLeuGlyIleGlnTrpArgLeuHisIleProCysTrpProGln 984  
QY 5154 AGCTCAGGACAGTAGAGATGAATAGAACCAATTAAGAGACCTTACTAATGACC 5213  
Db 985 ThrSerGlyLysValGluArgAlaAsnGlyIleLeuLysAlaGlnLeuThrLysLeuThr 1004  
QY 5214 GCGGAGACTGGCGTTAATGATTGATAGTCTCTGCTCCCTTTGTGCTTTTGGGTAGG 5273  
Db 1005 LeuGluVal---GlnLysProTrpThrSerLeuLeuProIleAlaLeuGluSerIleArg 1023

QY 5274 AACACCCCT---GGACAGTTTGGGCTCACCCCTTATGAATTTACTCTACGGGACCC--- 5327  
Db 1024 AlaSerProLysAlaProSerPheLeuSerProPheGluLeuIleTyrGlyArgProPhe 1043  
QY 5328 -----CCCCCATTTGGTAGAAATTTGCTTCTGTACATAGTGTGACGTGCTG 5372  
Db 1044 LeuLeuGlnAsnArgProPro-----SerAsnSerGlnLeuGlyGluTyrLeu 1059  
QY 5373 CTTTCCAGCCTTTGCTCTAGCTCAAGGCACCTTTCAGTGGGTGAGACACAGCGCTGG 5432  
Db 1060 ProThrValSerLeuMetSerTyrLeu-----LeuCysGlnGlnAlaAsp 1074  
QY 5433 AGCAAACTCCGGAGCCCTACTCAGGAGGAGGAGACTTTCAG-----ATC 5477  
Db 1075 GlnAlaLeuProLysProHisGluGlyValSerAsnProLys\*\*\*ThrCysSerProIle 1094  
QY 5478 CCACATCGTTTCCAACTGGGAGATTCA-----GTCTACGTTAGACCCACCGTGCA 5528  
Db 1095 ProLys-----AspSerLeuSerArgValThrLeuGlnAsnHisArgGly 1109  
QY 5529 GGAACCTCGAGACTCGGTGGAGGCGCTTATCTCTGTTACTTTG----- 5573  
Db 1110 LeuAspLeuLeuThrAlaGluLysGlyLeuCysIlePheLeuGluGluCysCys 1129  
QY 5574 -----ACCACACCAACCGCTGTG-----AAAAGTC 5597  
Db 1130 PheTyrThrAsnGlnSerGlyLeuValGlnAspAlaAlaGlyArgIleAsnGluLysAla 1149  
QY 5598 GAAGGAATCTCCACCTGGATCCATGTCATCCAGCTTAAACCGCGCCACCTCCCGATTCG 5657  
Db 1150 SerGlyArgValGlnTrpLeu-----ThrProValIleProAsp 1162  
QY 5658 GGTGTGAAAGCCGAAAG-----ACTGAAAAAT 5684  
Db 1163 LeuTrpGluAlaGluAlaGlyGlySerArgGlyGlnGluIleGluThrIleLeuAlaAsn 1182  
QY 5685 CCCCTTAAAGCTTCGCCTCCAT 5705  
Db 1183 ThrValLysProArgLeuTyr 1189

## RESULT 2

US-10-821-234-1083  
; Sequence 1083, Application US/10821234  
; Publication No. US20050255114A1  
; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Andarmani, Susan  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia  
; FILE REFERENCE: 821A  
; CURRENT APPLICATION NUMBER: US/10/821,234  
; CURRENT FILING DATE: 2004-04-07  
; PRIOR APPLICATION NUMBER: US 60/462,047  
; PRIOR FILING DATE: 2003-04-07  
; NUMBER OF SEQ ID NOS: 1704  
; SOFTWARE: pt\_seq\_genes Version 1.0  
; SEQ ID NO 1083  
; LENGTH: 1150  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(1150)  
; OTHER INFORMATION: Xaa = any amino acid or nothing  
US-10-821-234-1083

Alignment Scores:  
Pred. No.: 1,77e-107 Length: 1150  
Score: 1646.50 Matches: 438  
Percent Similarity: 51.9% Conservative: 186  
Best Local Similarity: 36.4% Mismatches: 470

Query Match:	11.2%	Indels:	109
DB:	6	Gaps:	37
US-10-723-552-3 (1-8132) x US-10-821-234-1083 (1-1150)			
QY	2226	GAGTCTCTGGTTGATATACCGGAGCGAAACATTTCAGTGTCTACTACAGCCATTAGGAAAACTA	2285
DB	4	GlnPheLeuValAspSerArgAlaAaCysSerValLeuThrGlnProIleCysLeuLeu	23
QY	2286	AAAGATAAAAATCCTGGGTGATGGGTGCCACAGGGCAACACAGTATCCATGGAGTACC	2345
DB	24	SerAsnCysAsnCysArgValMetGlyIle-----AspArgCysProLysValArg	40
QY	2346	CGAAGACAGCTGACTGGAGTGGGAGCGGTAAACCCAC-----TCGTTTCTG	2393
DB	41	LysPheThrPheProLeuAlaCysGluAlaMetSerArgLeuPheSerProHisPheLeu	60
QY	2394	GTCACTACTGAGTCCCGACAGCCCTCTTAGGTAGAGACTTATTAGCAAGATGGGAGCA	2453
DB	61	CysProGluCysProThrProLeuLeuGlyArgAspLeuLeuSerLysMetGlyAla	80
QY	2454	CAAAATTTCTTTGAACAAGGGAACACAGAAGTGTCTCAAAAT---AACAAACCTTACTCT	2510
DB	81	ThrIleSerLeuGluGluAspArgLeuGlnValGluAlaGluProGluGlnGlyIleHis	100
QY	2511	GTGTTGACCTCCAAATAGATGACGAATATCGACTATCTCTCCCTAGTAAAGCCGTAT	2570
DB	101	LeuLeuAlaLeuLeuAsnGlyGlnGluLeuGlu-----Thr	112
QY	2571	CAAAATATA---CAATCTCTGGTGGAAACAGTTTCCCAAGCTGGCGCAGAAACCGCAGGG	2627
DB	113	GlnAsnIleProLysGluSerLysAspHisIleThrProSerLeuSerAspThrSerVal	132
QY	2628	ATGGGTTTGGCAAGCTTCCCGACAAAGTTTCAACTGAAGGCGCATGGCCACCA	2687
DB	133	LeuGlyGlnAlaAsnLysValProProValLysThrAspLeuLysProGlyMetGlyTyr	152
QY	2688	GTGTCAGTCAGAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTCGGCCGCATGTC	2747
DB	153	ProTrpArgLysSerTyrLeuLeuLysProValAlaLeuAsnGlyValGlnProLeuLeu	172
QY	2748	CAAAAGATTAAATCCAAAGGGCATCTAGTTCCTGCTCAATCTCCCTGGAATACTCCCGTG	2807
DB	173	HisLysPheLeu---GlnGly---LeuArgProCysGlnSerProGlyAsnThrSerVal	191
QY	2808	CTACCGGTTAGAAAGCTGGGACTTAATGACTATACGACCATCGACGAGGCTTGAGAGGCT	2867
DB	192	LeuProValLysLysPro---AsnGlyGluTyrGlnPheVal***AspLeuLysGlyVal	210
QY	2868	AATAAACGGGTGACAGGATATACACCCACAGCTCCCGAACCTTATAACCTCTGTGTGCT	2927
DB	211	AsnGluAlaValIleProIleHisProThr---ProGlyPro---LysLeuValCysAla	228
QY	2928	---CTCCACCCCAACGGAGCTGGTATACAGTATTGGACTTAAAGGATGCTCTCTCTGC	2984
DB	229	GlnIleProGlyAspAlaGlnPhePheThrLeuLeuHisPheLysAspAlaPhePheCys	248
QY	2985	CTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTTCGAAATGGAGAGATCCAGGTACG	3044
DB	249	ValAlaLeuSerThrGlnAsnProCysIlePheCysPheGluTrpArgAsnProAspThr	268
QY	3045	GGAAGAACCAGGAGCTCACTGACCCGAGCTGCCCAAGGTTCAAGAACTCCCGACC	3104
DB	269	GlnLysAlaThr***TyrTrpThrValLeuProGlnGlyPheTrpGluGlyProHis	288
QY	3105	ATCTTTGACGAGCGCTTACACAGAGACTGGCCCAACTTCAGGATCCAAACCCCTCAGGTG	3164
DB	289	LeuPheGlyAsnAla-----GlySer***AlaTyrArgMetGly-----	301
QY	3165	ACCTCTCTCAGTACGTGGATGACTGCTGCTGCGGGAGCGCCCAACAGGACTGCTTA	3224
DB	302	LeuLeuLeuGlnTyrValAspAspLeuPheIleValSerGlnThrTrpGlnAspSerAsp	321

QY	3225	GAAGGCACGAAGGCACTACTACTGCTGGAATTTGCTGAGCTAGGCTACAGAGCCTCTCTAAG	3284
DB	322	PheAsnIleIleLysThrLeuAsnPheLeuAlaGluArg***TyrLysAlaAlaProSer	341
QY	3285	AAGCCCCAGATTTTGCAGGAGAGAGGTAACATACTTTGGGGTACAGTTTGGCGGACGGGACG	3344
DB	342	LysAlaGlnIleSerLeuGlnLysPheArgTyrLeuGlyPheIleLeuThrProGlyAla	361
QY	3345	CGATGGCTGACGAGGACGGAAGAAAACCTGTAGTCAGATACCGGCCCAACACACAGCC	3404
DB	362	ArgThrLeuValAspGlyLeuGluLysAlaIleThrSerLeuLeuValProGlnThrLys	381
QY	3405	AAACAATGAGAGAGTTTGGGGACAGCTGGATTTTGCAGCTGTGGATCCCGGGGTTT	3464
DB	382	ArgGlnProGlnGlyPheLeuGlyMetValGlyPheCysSerIleTrpIleProAsnTyr	401
QY	3465	CGACCTTAGCAGCCCACTCTATCCCGCTAAACCAAA-----GAAAAGAGGGAATTCTCC	3518
DB	402	GlyLeuThrAlaLysProLeuCysGluThrGlnLysGlyGluArgGluProLeuTyr	421
QY	3519	TGGGCTCTGACGACACGAGAGGATTTGATGCTATCAAAAAGGCCCTGTGAGCGCACCT	3578
DB	422	TrpGluLysLysCysGlnLeuSerPheGluProLeuLysThrGluLeuGlyGlnAlaLeu	441
QY	3579	GCTCTGCGCTCCCTGACGTAACTAAACCTTTTACCTTTATGTGATGAGCGCTAAGGGA	3638
DB	442	ValLeuGlyLeuProAsp---GluLysProLeuThrLeuTyrValargGluArgLeuGly	460
QY	3639	GTAGCCCGGGAGCTTTTAAACCCAAACCTTAGGACCATGGAGAGACTGTGCGCTACCTG	3698
DB	461	IleAlaLeuGlyValLeuThrGlnArgLeuGlyProValGlnLysProValAlaHisPhe	480
QY	3699	TCAAAGAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCTGCTGAAGGCTATCGAGCT	3758
DB	481	SerLysGlnThrAsnProValAlaGlnLysArgProGlyCysTrpArgAlaValAlaAla	500
QY	3759	GTGCGCACTACTGTCAAGGACGCTGACAAATTCACTTTGGGACAGAAATATAACTGTATA	3818
DB	501	ThrAlaLeuLeuPheSerGluAlaSerLysLeuThrGlnGlyGlnTyrThrGluAlaMet	520
QY	3819	GCCTCCCATGTCATGGTAGAACATCGTTTGGGACGCCCCCAGACCGATGGATGACCAACGCC	3878
DB	521	ThrProHisGlnValGlnThrGlyLeuGluValLysGlyHisIleTrpLeuThrAlaGly	540
QY	3879	CGATGACCCACTATCAAGCCCTGCTTCTC---ACAGAGAGGCTCAGCTTCGCTCCACCA	3935
DB	541	ArgLeuThrAsnTyrGlnAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu	560
QY	3936	GCCGCTCTCAACCTGCCACTCTTCTGCTGGAAGAGACTCATCAACCAAGTACTCATGAT	3995
DB	561	GlnThrLeuAsnThrAlaThrSerLeuProThrThrAspGluAspSerGlnHisGln	580
QY	3996	TGCCATCAACTATTATTGAGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTG	4055
DB	581	CysThrGly---ProThrLysGlnThrHisSerArgProSerLeu***AspLysProLeu	599
QY	4056	ACTGGAGAGTGTCTACCTGGTTCACTGACGGAAGACGCTATGTGTGGAGAGTAAGAGG	4115
DB	600	GluAsnAlaLysThrGlu***TyrProAspGlySerSerPhe---ValGluGlyThrGln	618
QY	4116	ATGCTGGGCGCGGTGGTGGACGGGACCCGACGATCTGGGCCAGCAGCTGCCGGAA	4175
DB	619	LysAlaGlyHisAlaValValSerLeuLysGluThrThrGluThrLysAlaLeuProPro	638
QY	4176	GGAATTCAGCACAAAAGCTGAGCTCATGCCCTCAGCAAGCTTTGCGGCTGGCCGAA	4235
DB	639	GlnThrSerAlaGlnLysAlaAlaLeuGluAlaLeuThrGluProHisAsnTrpGluLys	658
QY	4236	GGAATTCATAACATTTATACGACAGCAGGTATCCCTTTGGACTGCACACGTACAT	4295
DB	659	GlyArgArgValAsnLeuTyrThrAsnSerLysTyrArgPheLeuLeuLeuCysAlaHis	678
QY	4296	GCGGCCATCTATAAACAAGGGGGTTCCTTACCTCAGCAGGAGGGGAAATAAAGACAAA	4355



```

Db      679 AlaAlaIleTrpLysGluLysGlyMetLeuThrAspLysAsnSerProGlyLysArgLys 698
QY      4356 GAGGAAATCTTAAAGCCATTAGAACCGGTACATTTACCAAAAGAGTAGCTATTATATACAC 4415
Db      699 AspProIleGlnLeuLeuGluAlaGlnLeuProCysGlnValAlaValIleHis 718
QY      4416 TGTCTCGACATCAGAAAGCTTAAAGATCTCATATCCAGAGAAACACAGATGGCTGACCGG 4475
Db      719 CysLysGlyHisGlnArgAspTrpSerLeuIleSerGlnGlyAsn-----Arg 734
QY      4476 GTTGCCCAAGCAGGCA-----GCCAGGGTGTAAACCTTCCTGCTCTATAATAAGAATG 4526
Db      735 AlaThrLysGlnAlaThrTrpLeu**GlusSerAsnValLeuAlaLeuMetThrAsp 754
QY      4527 CCCAAAGCCCGAAGACCCAGACGACGATACACCCCTAGAACGTCGCGAAGAGATAAAAG 4586
Db      755 ProThrAsnLeuProAsnHisLeuGlnTrpSerGlnGlnGlu---GlnGlu**AlaLys 773
QY      4587 ATAGACAGATTCTGTAGACTCCGGAAGGACCTGTATACCTCAGATGGGAAGGAATC 4646
Db      774 GlyGlnGlyTrpGlnGluAsnSerAlaGly-----TrpPheThrAspGlyGluSerPhe 791
QY      4647 CTGCCCCACAAGAGGGTTAGATATGTCCACAGATACATCGTCTAACCCACCTAGGA 4706
Db      792 IleProGlyThrAspGlnLeuLysValThrLysHisLeuHisAspAlaThrHis----- 809
QY      4707 ACTAAACACCTGCAGCAGATGGTGCAGAACATCCCTTATCATGTCTTCGAGGCTACCAGGA 4766
Db      810 -----**TrpArgAspValLeuTrpAsnSerValHisLysIlePheThrArgLysLys 827
QY      4767 GTGGCTGACTCGTGTCAACATATTGTGCCCTGCAGCTGGTTAATGTAATCTCTCC 4826
Db      828 LeuArgLysSerValLysGlnValThrPheValSerGluLeuCysSerArgAsnPro 847
QY      4827 AGA-----ATGCTCCAGGGAAG-----AGACTAAGGGGAAGCCACCCA 4865
Db      848 HisThrHisProIleProSer**LeuLysLeuValGlnHisArgGlyThrTrpLeu 867
QY      4866 GCGCTCCTCCTGGAAGTGGACTTCACTGAGTAAAGCGGCTAAATACGGAACAATAC 4925
Db      868 GlyGluAspTrpGlnValValLeuThrGlnMetThrPro---AsnLeuGlyTrpLysTrp 886
QY      4926 CTATGGTTTTGTAGACACTTTTTCAGGATGGGTAGAGCTTATCTACTAAGAAGAG 4985
Db      887 LeuLeuValPheValAspThrPheThrGlyArgValLysGlyPheProThrCysThrGlu 906
QY      4986 ACTTCAACCTGGTGGCTTAAATAATCTCGAAGAAATTTTTCAGATTTGGAATACCT 5045
Db      907 LysAlaValGluValCysLysProSerLysGluValIleSerGlnPheGlyLeuPro 926
QY      5046 AAGTAATAGGTGCAGACATGTGCCAGCTTTTGTGCCCGAGTAAAGTCAAGGAGCTGGCC 5105
Db      927 LysSerProGlnSerGlyAsnArgLysLeuSerPheMetGlyLysIleThrGlnSerLeuSer 946
QY      5106 AAGATATTGGGGATTGATTGGAAACTGCATTTGTGCATACAGACCCCAAGACTCAGACAG 5165
Db      947 ThrThrLeuGlyIleAspTrpGluLeu-----AlaProGlnSerSerGlyLys 962
QY      5166 GTAGAGAGGATGATAGAACCATTAAGACACCTTACTAAATTTGACCGCGGAGACTGGC 5225
Db      963 Val---LysMetAsnHisThrLeuGluThrThrLeuAlaLysLeuPheGlnGluMetHis 981
QY      5226 GTTAATGATGGATAGTCTCTCCCTGCTTTTGTGCTTTTATAGGCTTAGGAACCCCTGGA 5285
Db      982 ---GluSerTrpValLysMetLeu---LeuSerLeuLeuArgValArgAlaProSer 999
QY      5286 ---CAGTTTGGGCTGACCCCTATGAATTAATCTACGGGGGAGCCCCCATTTGGTAGAA 5342
Db      1000 CysSerLeuArgLeuSerSerGlnMetIleTrpGlnArgPro----- 1014
QY      5343 ATTGCTTCTGTACATAGTGTGACGTGCTTCTCCAGGCTTTTGTCTCTAGGCTCAAG 5402

```

```

Db      1015 -----PheLeuThrThrAspLeuLeuGlyGluLeuHisLysLysLeu--- 1030
QY      5403 GCACCTTGGTGGTGAGACAAACGAGCGTGAGGCAACTCCGGGAGGCTACTCAGGAGGA 5462
Db      1031 -----GlnTyArgPheArgThrAspSerLys 1039
QY      5463 GGAGACTTGCAGATC-----CCACATCGT----- 5486
Db      1040 GlyAspPheArgIleTrpLysGlnAsnProAlaSerProHis**GlyTrpGluAsnPro 1059
QY      5487 TTCCAAGTG-----GGAGATTACAGTCTAGTACGCCACCGTCGAGGAACCTCGAG 5540
Db      1060 PheGlnValAsnProGlyAspGlnValLeuLeuLysAla***ArgAlaGlySerProGlu 1079
QY      5541 -----ACTCGTGAAGGCGCTTATCTGCTACTTTTGACACACCAACCGCT 5588
Db      1080 AspHisProLeuLeuLysTrpGluGlyProCysTrpValIleLeuThrThrProThrAla 1099
QY      5589 GTCAAAAGTCGAAGATCTCCACCTGGATCCATGTCATCCACGTTAAACCGGCGCCACCT 5648
Db      1100 AlaAsnSerGlnGlyIleThrSerTrpValHisLeuSerArgSerGluMetLeuSerPro 1119
QY      5649 -----CCGATTCCGGTGGAAAGCCGAAAGACTGAAATCCCTT 5690
Db      1120 LysCysPheGlnThrArgProAspGlyProPheTySer***LysProValGluAspLeu 1139
QY      5691 AAGCTTCGC 5699
Db      1140 LysPheArg 1142

RESULT 3
US-11-129-442-47
; Sequence 47, Application US/11129442
; Publication No. US20060002951A1
; GENERAL INFORMATION:
; APPLICANT: Kleiman, Lawrence
; APPLICANT: CEN, Shan
; APPLICANT: GUO, Fei
; TITLE OF INVENTION: Inhibition of the tRNAlys3-Primed Initiation of Reverse
; FILE REFERENCE: 11168.257
; CURRENT APPLICATION NUMBER: US/11/129,442
; CURRENT FILING DATE: 2005-05-16
; PRIOR APPLICATION NUMBER: CA 2,467,312
; PRIOR FILING DATE: 2004-05-14
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 47
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Murine leukemia virus
US-11-129-442-47

Alignment Scores:
Pred. No.: 2,05e-87 Length: 537
Score: 1358.00 Matches: 277
Percent Similarity: 62.7% Conservative: 76
Best Local Similarity: 49.2% Mismatches: 142
Query Match: 9.3% Indels: 68
DB: 12 Gaps: 12

US-10-723-552-3 (1-8132) x US-11-129-442-47 (1-537)
QY      585 ATGGGACACACCGTGCAGCCCTCTTAGTTGACTCTCGACCATTTGGACTGAAGTTAA 644
Db      1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGluHisTrpGlyAspValGln 20
QY      645 TCCAGGCTCATAAATTTGTCAGTTTAAAGAGGACCTTTGGCAGACTTTCTCTGTC 704
Db      21 ArgIleAlaSerAsnGlnSerValAspValLysArgArgTrpValThrPheCysSer 40
QY      705 TCTGAATGCCGACATTCGATGTGGATCGCCATCAGAGGGACCTTTTAACTCTGAGATT 764

```





```
; SEQ ID NO 12
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-11-042-988-12

Alignment Scores:
Pred. No.: 1,49e-22 Length: 912
Score: 431.00 Matches: 209
Percent Similarity: 35.9% Conservative: 145
Best Local Similarity: 21.2% Mismatches: 345
Query Match: 2.9% Indels: 286
DB: 7 Gaps: 39

US-10-723-552-3 (1-8132) x US-11-042-988-12 (1-912)
QY 2391 CTGTCATACCTGAGTCCCGCAGCAGCCCTCTTAGTAGAGACTTATTGACCAAGATGGGA 2450
D 40 ValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGly 59
QY 2451 GCACAAATTTCTTTGAAACAGGGAACAGAGTGTCTGCAATAACAACCTATCACT 2510
D 60 CysThrLeuAsnPhe-----ProIle--- 66
QY 2511 GTGTTGACCTCCAAATTAGATGACGAATATCGACTATATCTCCCTAGTAAAGCCTGAT 2570
D 67 -----SerProIle----- 69
QY 2571 CAAATATACAAATCTGTTGGAACAGTTCCTCCCAAGCCTGGGCAGAAACCGCAGGATG 2630
D 69 ----- 69
QY 2631 GGTGTGGCAAGAGTTCCCCCAAGATTATTCACCTGAAGCCAGTGCACACCAAGTG 2690
D 70 -----GluThrValPro-----ValIysLeuIysProGlyMetAspGlyPro 83
QY 2691 TCAGTCAGACAGTACCCTTGAGTAAAGAGCTCAAGAGGAATTCGGCCGCATGTCCAA 2750
D 84 LysValIysGlnIrrProLeuThrGluGluIysIleIysAlaLeuValGluIleCysThr 103
QY 2751 AGATTAAATCCCAAGGCGATCTAGTTCCTGTC-----CAATCCCTCGGAATPACTCCC 2804
D 104 GluMetGluIysGluGlyIleSerLysIleGlyProGluAsnProTyrAsnThrPro 123
QY 2805 CTGCTACCGGTGAAAGCCTGGGACTAATGACTATCGACAGTACAGGACTTGAGAGAG 2864
D 124 ValPheAlaIleLysLysAspSerThrIysTrpArgLysLeuValAspPheArgGlu 143
QY 2865 GTCAATAAAGGGTGCAGGAT-----ATACACCAACAGTCCCGAACCTTATAAC 2915
D 144 LeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGly 163
QY 2916 CTCCTGTGTGCTCCCAACCGAGCTGGTATACAGTATTGGACTTAAAGGATGCC 2975
D 164 Leu-----LysLysLysSerValThrValLeuAspValGlyAspAla 178
QY 2976 TTCCTTCCTCGCTGAGATTACCCCACTAGCCCAACCTTTTTCCTTCGATGAGAGAT 3035
D 179 TyrPheSerValProLeuAspGluAspPheArgLysTrpThrAlaPheThrIleProSer 198
QY 3036 CCAGGTACGGGAAGAACCGGG--CAGTCACCTGGACCGCAGTCCCAAGGTTTCAAG 3092
D 199 IleAsnAsnGluThrProGlyIleArgTrpGlnIrrAsnValLeuProGlnGlyTrpLys 218
QY 3093 AACTCCCGCGACATCTTTGACGAAGCCCTACACAGACCTGGCCAACTTCAGGATCCAA 3152
D 219 GlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGln 238
QY 3153 CACCTCAGGTGACCTCTCTCAGTACGTGATGACCTGCTTCTGGGGAGGCCACCAAA 3212
D 239 AsnProAspIleValIleTyrGlnTyrMetAspAspLeuVal---GlySerAspLeu 257
QY 3213 CAGGACTGCTTAGAAGGCACCAAG-----GCACCTACTGCTGGAATTTGCTGACCTAGGC 3266
```

```

258 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuLeuArgTrpGly 277
3267 TACAGAGCCTCTCTAAGAAGGCCAGATTTCGAGAGAGAGAGTAACATACTTG----- 3320
278 LeuThrThrProAspLysLysHisGln-----LysGluProPheLeuTrpMet 294
3321 GGGTACAGTTTGGGACGGCAGCGATGCTGACGAGGACCGGACCGAAGAAACCTGTAGTC 3380
295 GlyTyrGluLeuHisPro---AspLysTrp-----ThrValGln 306
3381 CAGATACCGGCCCAAC-----ACAGCCAAACAAATAGAGAGATTTTGGGG 3428
307 ProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 326
3429 ACAGTGGATTTCAGACGTGTGGATCCCGGGTTT----- 3464
327 LysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeu 346
3465 -----CGACCTTAGCAGCCCTCTACCGCTAACCAA----- 3500
347 LeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGlu 366
3501 -----GAAAGGGGAATTCCTCGGCTCTCTGAGCAGCAGGCAAGCAATTTGATGCTATC 3554
367 LeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSer 386
3555 AAAAAAGCCCTGCTGAGCGCAGCTCTCTGCGCCCTCTCTGAGTAACTAAACCTTTACC 3614
387 LysAspLeuIleAlaGluIleGlnLysGlnGlyGlnIrrThr-----TyrGln 404
3615 CTTTATCTGATGAGCGTAGGAGTAGCCCGGGGAGTTTAAACCAACCTAGGACCA 3674
405 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 424
3675 TGAGAGAGACCTCTGCGCTACCTGTCAAAAGCTGATCTCTGAGCAGTGGTGGGCC 3734
425 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr----- 441
3735 ATATGCTGAAGCTATCGCAGCTGGCCATCTGTCAGGACGCTGACAAATTCAGT 3794
442 -----GluSerIleValIleTrpGlyLys----- 449
3795 TTGGACAGAATAATCTGTAATAGCCCCCATGCTGAGGAACATCGTTGCGCAGCCC 3854
450 -----ThrProLysPheLysLeuProIleGlnLysGluThr 461
3855 CCAGACCGATGATGACCAACGCCGATGACCCACTATCAAGAGCCTGCTTCTCACAGAG 3914
462 TrpGluThrTrpTrpThrGlu-----TyrTrpGln----- 471
3915 AGGTACAGTTCGCTCCACCGCGCTCTCAACCTGCCACTCTTCTGCTCGAA----- 3968
472 -----AlaThrTrpIleProGluTrpGlu 479
3969 ---GAGACTGATGAACGAGTACTCATGTTGTCATCACTATTTGATTGAGGAGACTGGG 4025
480 PheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGlu----- 496
4026 GTCCGCAAGGACCTTACAGACATACCGCTGACTGAGGAGAGTGTAACTGGTTCACTGAC 4085
497 -----ProIleValGly---AlaGluThrPheTyrValAsp 507
4086 GGAAGCAGC-----TATGTGGTGGAGGTAAAGG 4115
508 GlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArg 527
4116 ATGCTGGGGCGGCTGGGAGCGGACCGCAGCATCTGGGCGCAGCAGCTGCCCGAA 4175
528 GlnLysValValThrLeuThrAsp----- 535
4176 GGAACCTTACACAAAAGGCTGAGCTCATGCCCTCAGCAAGCTTTGCGCTGCCGAA 4235
```

D	536	---	Thr	Thr	Asn	Gln	Leu	Thr	Glu	Leu	Gln	Ala	Leu	Leu	Gln	---	Asp	Ser	553
Q	4236	GGG	AAAT	CCATAA	CAATTTAT	ACG	GAC	GAC	GAG	TAT	GCTTTG	CGACT	GCAC	AC	GTA	CA	TACAT	4295	
D	554	Gly	Leu	Glu	Val	Asn	Ile	Val	Thr	Asp	Ser	Gln	Tyr	Ala	Leu	Gly	---	568	
Q	4296	GGG	CCAT	CTATAA	CAAA	GGGG	TGTT	GCTT	AC	G	CG	AGG	AGG	GA	AA	TA	AA	4355	
D	569	---	Ile	Ile	Gln	Ala	Gln	Pro	Asp	Gln	Ser	Glu	Ser	579					
Q	4356	GAG	AAAT	TCTAAG	CCCTATT	AGA	AGC	CGCTAC	ATT	TAC	CAAA	AGG	CTAG	CTAT	TAT	AC	4415		
D	580	Glu	Leu	Val	Asn	Gln	Ile	Ile	Glu	Gln	Leu	Ile	Leu	Leu	Val	Tyr	Leu	599	
Q	4416	TGT	CTG	CACAT	CAGA	AGCTAA	AGAT	CTCAT	TAT	CC	AG	AGAA	ACC	AGAT	GGCTG	AC	CGG	4475	
D	600	Val	Pro	Ala	His	Leu	Ser	Gly	Ile	---	---	---	Gly	Gly	Asn	Glu	Val	614	
Q	4476	GTT	CCA	GAC	GGC	AGC	CGC	AGG	TGTTAA	CCCTT	CGC	CTAT	ATA	TAGA	AA	TG	CC	4535	
D	615	Leu	Val	Ser	Ala	Gly	Ile	Arg	Gly	Val	Leu	Phe	Leu	---	---	---	---	626	
Q	4536	CC	AGA	ACC	AG	AC	GAC	AGT	CAC	CC	TAGA	AG	ACT	GG	CA	AG	AGAT	4595	
D	626	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	626	
Q	4596	TT	CT	CTG	AG	CT	CCG	GA	GG	AC	CTG	CTAT	AT	CT	CT	CA	T	4655	
D	626	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	626	
Q	4656	AA	GA	AGG	TT	PAGA	ATAT	GT	CC	AA	CAG	ATAC	TG	CT	TA	ACC	CA	4715	
D	627	---	Asp	Gly	Ile	Asp	Leu	Ala	Gln	Asp	Glu	His	Glu	Leu	Ser	Thr	Arg	645	
Q	4716	CT	G	AG	CAG	AT	GGT	TC	AGA	CA	TCC	CTT	AT	C	AG	TT	CT	4775	
D	646	Met	---	---	---	---	---	---	Ala	Ser	Asp	Phe	Asn	Leu	---	Pro	Val	658	
Q	4776	TC	G	TG	T	CA	AA	CA	T	TG	T	GC	CT	CG	CA	CT	G	4835	
D	659	Glu	Ile	Val	Ala	Ser	Cys	Asp	Leu	Cys	Gln	Leu	---	---	---	---	---	669	
Q	4836	CC	AG	GA	G	AG	ACT	AA	AGG	GA	GC	---	---	---	CA	CC	CA	4886	
D	670	Lys	Gly	Glu	Ala	Met	His	Gly	Gln	Val	Asp	Cys	Ser	Pro	Gly	Ile	---	688	
Q	4887	TT	CA	CT	CAG	TA	AG	CC	GG	CTAA	AT	AC	G	GA	A	CA	AT	4946	
D	689	Cys	Thr	His	Leu	Glu	---	---	---	---	---	Gly	Lys	Val	Ile	Leu	Val	703	
Q	4947	TT	TT	CA	G	TG	GGT	PAG	AGG	CTTAT	CT	CT	TA	AG	AA	AG	AG	5006	
D	704	Ala	Ser	Gly	Tyr	Ile	Glu	Ala	Glu	Val	Ile	Pro	Ala	Glu	Thr	Gln	Ala	723	
Q	5007	AAA	TA	CT	CG	GA	AG	AAAT	TTTT	TCC	AA	GAT	TTG	GA	T	AC	CT	5066	
D	724	Phe	Leu	Leu	---	Lys	Leu	Ala	Gly	Arg	Trp	Pro	Val	---	Lys	Thr	Ile	741	
Q	5067	GG	T	CA	CA	CT	TT	TG	T	GT	CC	CA	GGT	AA	GT	CA	GG	5126	
D	742	Gly	Ser	Asn	Phe	Thr	Gly	Ala	Thr	Val	Arg	Ala	Ala	Cys	Trp	Trp	Ala	761	
Q	5127	AAA	CT	CA	T	GT	TG	CA	CA	AC	CC	AA	AG	CT	CA	GG	T	5186	
D	762	Glu	Phe	Gly	Ile	Pro	Tyr	Asn	Pro										

QY	6547	GCTTCGGC-----	CCACCTTACTATGAAGATGGCT	6579
RESULT 6				
US-10-821-234-1395				
US-10-821-234-1395				
Sequence 1395, Application US/10821234				
Db	193	ProLeuAsnPheArgProTyrValSerIleProValProGluGlnTyrAsnAsnPheSer		:::

```
QY 6580 AGAAGA-----GGGAATTCATGTGACA 6603
Db 213 ThrGluLeuAsnThrThrSerValLeuValGlyProLeuValSerAsnLeuGluLeuThr 232
QY 6604 AAAGAACATAGAGACCAATGCATCGGGATCCCAAAATAGCTTACCCTTACTGAGGTT 6663
Db 233 HisThrSerAsnLeuThrCysVal-----LysPheSerAsnThrThrTyr 247
QY 6664 TCTGGAAAGGCACCTGCATAGGAAGAGTCCCCCATCCACCAACACCTTTGTAAACCAC 6723
Db 248 ThrThrAsnSerGlnCysIleArgTrpValThrProThrGlnIleValCysLeuPro 267
QY 6724 ACTGAAGCCCTTAATCAAACTCTGAGAGTCAATATCTGCTACCTGCTTATGACAGGTGG 6783
Db 268 SerGlyIlePhe----- 271
QY 6784 TGGGCATGTAATACTGATTAACCCCTTGTTTCCACCTTGGTTTTTAACCAAACTAAA 6843
Db 272 PheValCysGlyThrSerAlaTyrArgCysLeuAsn-----GlySerSerGlu 287
QY 6844 GATTTTGGATATATGTCCTCAAAATGTTCCCGAGTGATTAATATCCCGAAAAAGCAATC 6903
Db 288 SerMetCysPheLeuSerPheLeuValProProMetThrIleTyrThrGluGln----- 305
QY 6904 CTTGATGAATATGACTAC-----AGAAATCATCGACAAAGAGAGAACCCATATCTCTG 6957
Db 306 ---AspLeuTyrSerTyrValIleSerLysPheProArgAsnLysArgValProIle----- 322
QY 6958 ACACCTTGCTGTATGCTCGGACTTGAGTGGCAGAGGTGTAGGAAACAGGAACAGCTGCC 7017
Db 323 ---LeuProPheValIleGlyAlaGlyValLeuGlyAlaLeuGlyThrGlyIleGlyGly 341
QY 7018 CTGCTACAGGGACACAGAGCTAGAAACAGAGCTTAGTAACCTACATCGAATTTGTAACA 7077
Db 342 IleThrThrSerThrGln-----PheTyrTyrLysLeuSerGlnGluLeuAsn 357
QY 7078 GAAGATCTCCAAAGCCCTAGAAAATCTGTCAGTAACCTGAGGAGNATCCCTAACTCTTA 7137
Db 358 GlyAspMetGluArgValAlaAspSerLeuValThrLeuGlnAspGlnLeuAsnSerLeu 377
QY 7138 TCTGAAGTAGCTCTACAGAAATAGAGAGGTGTAGATTATTATTCTTAAAGAGGAGGA 7197
Db 378 AlaAlaValValLeuGlnAsnArgArgAlaLeuAspLeuLeuThrAlaGluArgGlyGly 397
QY 7198 TTATGTACCTCTGAAGGAGGAATGCTGTTTTTATGTGATCATTCAGGGGCCATCAGA 7257
Db 398 ThrCysLeuPheLeuGlyGluGluCysCysTyrTyrValAsnGlnSerGlyIleValThr 417
QY 7258 GACTCCATGAACAGCTTAGAGAAAGGTTCGAGAGCGTCGAGAGGGAAGAAACTACT 7317
Db 418 GluLysValLysGluIleArgAspArgIleGlnArgArgAlaGluGluLeuArgAsnThr 437
QY 7318 CAAGGCTGGTTGAGGATGGTTCAACAGGTCTCTTTGGTTGGCTACCTTCTTGCT 7377
Db 438 GlyProTrp-----GlyLeuLeu-----SerGlnTrpMetProTrpIleLeuProPhe 453
QY 7378 TTAACAGGACCTTAATAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7437
Db 454 Leu---GlyProLeuAlaIleIleLeuLeuLeuPheGlyProCysIlePheAsn 472
QY 7438 AAGTTAATTCCTTATTAGAGAAAGCAATAGTGCAGTCCAGATC---ATGGTACTTAGA 7494
Db 473 LeuLeuValAsnPheValSerSerArgIleGluAlaValLysLeuGlnMetGluProLys 492
QY 7495 CAACAG-----TACCAAAGCCCGTCTAGCAGGGAAGCT 7527
Db 493 MetGlnSerLysThrLysIleTyrArgArgProLeuAspArgProAla 508
```

RESULT 7

US-10-507-928-10

; Sequence 10, Application US/10507928

; Publication No. US20050266024A1

```
; GENERAL INFORMATION:
; APPLICANT: POWDERMED LIMITED AND GLAXO GROUP LIMITED
; TITLE OF INVENTION: ADJUVANT
; FILE REFERENCE: N.89232B GCW
; CURRENT APPLICATION NUMBER: US/10/507,928
; CURRENT FILING DATE: 2004-09-17
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of RT insert of p7077-RT3
US-10-507-928-10
```

```
Alignment Scores:
Pred. No.: 1,56e-13 Length: 562
Score: 301.00 Matches: 145
Percent Similarity: 37.2% Conservative: 104
Best Local Similarity: 21.6% Mismatches: 267
Query Match: 2.1% Indels: 154
DB: 6 Gaps: 26
```

US-10-723-552-3 (1-8132) x US-10-507-928-10 (1-562)

```
QY 2628 ATGGGTTTGGCAAGCAAGTTCCCCACAAAGTTATTCAACTGAAGGCCAGTGCACACCA 2687
Db 1 MetGlyProIleSerProIleGluThrValSerValLysLeuLysProGlyMetAspGly 20
QY 2688 GTCTCAGTCAGACAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTCGGCCGCGATGC 2747
Db 21 ProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCys 40
QY 2748 CAAAGATTAAATCCACAGGCGCATCTAGATTCTCTGTC-----CAATCTCCCTGGAATPACT 2801
Db 41 ThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThr 60
QY 2802 CCCTGCTACCGTTAGAAAGCTGGGACTAATGACTATGACCATGACAGGACTAGAGGACTTGAGA 2861
Db 61 ProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 80
QY 2862 GAGGTCAATAAAGCGGTGCAGGAT-----ATACACCAACAGTCCCGAACCCCTTAT 2912
Db 81 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 100
QY 2913 AACCTCTGTGCTCTCCACACCCCAACGAGAGTGTGTATACAGTATTGGACTTAAAGGAT 2972
Db 101 GlyLeu-----LysLysLysLysSerValThrValLeuAspValGlyAsp 115
QY 2973 GCCTTCTTCTGCTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTTGCCTTGCATGGAGA 3032
Db 116 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 135
QY 3033 GATCCAGGTACGGGAAGAACCGGG---CAGCTCACCTGACCCGACTGCCCAAGGGTTC 3089
Db 136 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 155
QY 3090 AGAACTCCCGACCATCTTTGAGAAAGCCCTACACAGAGACTGCGCAACTTCAGATC 3149
Db 156 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 175
QY 3150 CAACACCTCAGGTGACCTCTCCAGTAGTACGTGATGACTGCTTCTGGCGGAGGCCACC 3209
Db 176 GlnAsnProAspIleValIleTyrGlnTyrMetAspLeuTyrVal---GlySerAsp 194
QY 3210 AAACAGACTGCTTAGAAGGCACGAAG-----GCACACTGCTGGAATTTGCTGACCTA 3263
Db 195 LeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrp 214
QY 3264 GGCTACAGAGCCTCTCTCTAAGAGGCCCGCCAGATTTGCGAGGAGAGAGGTAAACATCTTG--- 3320
Db 215 GlyLeuThrThrProAspLysLysHisGln-----LysGluProProPheLeuTrp 231
```

3321 ---GGGTACAGTTTGGGGAGCGGACGCGATGGCTGACGGAGGACCGAAGAAACTGTA 3377  
Db MetGlyTyrGluLeuHisPro---AspLysTrp-----ThrVal 243  
3378 GTCCAGATCCGCCCCCAACC-----ACAGCCAAACAATGAGAGAGTTTGG 3425  
Db GlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuVal 263  
3426 GGCACAGCTGGATTTTGCAGACTGTGGATCCCGGGTTT----- 3464  
Db GlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLys 283  
3465 -----GCGACCTTAGCGCCCACTCTACCCGCTAACCAAA----- 3500  
Db LeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeu 303  
3501 -----GAAAAAGGGNAATCTCTGGCTCTGAGCACCAGAGGCAATTTGATGCT 3551  
Db GluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAspPro 323  
3552 ATCAAAAGCCCTGTGAGCGCACCTGCTCTGCGCTCCCTGACGTAACTAAACCTTT 3611  
Db SerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTrpThr-----Tyr 341  
3612 ACCCTTATGTGATGAGCGTAAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTTAGGA 3671  
Db GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 361  
3672 CCATGAGAGAGCTGTGCTACCTACCTGCTCAAGAAGCTGATCTCTAGCAGTGTGG 3731  
Db AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThr----- 379  
3732 CCCATATGCTGAAGCTATCGCAGCTGTGGCATACTGTCAAGCAGCGTGAACAATTG 3791  
Db GluSerIleValIleTrpGlyLys----- 387  
3792 ACTTTGGGACAGATATACTGTAATAGCCCCCATGTCATTTGGAGAACATCGTTCCGCGAG 3851  
Db ThrProLysPheLysLeuProIleGlnLysGlu 398  
3852 CCCCCAGCGATGGATGACCAACGCCCGCATGACCACCTATCAAGACCTGCTTCTCACA 3911  
Db ThrTrpGluThrTrpTrpThrGlu-----TyrTrpGln----- 409  
3912 GAGAGGCTCAGCTTCGCTCCACCGCGCTCTCAACCTGCGCACTCTCTGCTGAA--- 3968  
Db AlaThrTrpIleProGluTrp 416  
3969 -----GAGACTGATGAACAGTGAATCATGATGTCATCAATATTTGATTGAGGAGACT 4022  
Db GluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGlu----- 434  
4023 GGGGTCGCGAGACCTTACAGACATACGCTGATGAGAGAGTGTCACTGCTGTTCACT 4082  
Db ProIleValGly-----AlaGluThrPheTyrVal 444  
4083 GACGGAAGCAGC-----TATGTGTGGAGTAA 4112  
Db AspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGly 464  
4113 AGGATGGCTGGGCGCGGTGTGGACGGACCGCAGCATCTGGGCCAGCGCTCGCG 4172  
Db ArgGlnLysValValThrLeuThrAsp----- 473  
4173 GAAGGACTTACGACAAAGGCTGAGCTCATGGCCCTCAGCAAGCTTTGGCGTGGCC 4232  
Db ThrTrpAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGln---Asp 490  
4233 GAAGGGAATCCATAACATTTATACGGACAGCAGGTATGCTTTGCGACTGCACAGTA 4292  
Db SerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGly----- 506

4293 CATGGGCCCATCTATAAACAAAGGGGTTGCTTACTTCAGCAGGAGGGAATAAAGAAC 4352  
Db IleileGlnAlaGlnProAspGlnSerGlu 516  
4353 AAAGAGAAATCTAAGCCATTATTAGAGCGGTACATTACCAAAAAGGCTAGCTATTATA 4412  
Db SerGluLeuValAsnGlnIleileGluGlnLeuIleLysLysGluLysValTyrLeuAla 536  
4413 CACTGTCTCGGACATCAGAAAGCTAAAGATCTCATATCCAGAGAACACAGATGGCTGAC 4472  
Db TrpValProAlaHisLysGlyIle-----GlyGlyAsnGluGlnValAsp 551  
4473 CGGTTGCCAAGCAGCAGCCAGCGGTGT 4502  
Db LysLeuValSerAlaGlyIleArgLysVal 561

## RESULT 8

US-10-507-928-12  
; Sequence 12, Application US/10507928  
; Publication No. US2005026024A1  
; GENERAL INFORMATION:  
; APPLICANT: POWDERMED LIMITED AND GLAXO GROUP LIMITED  
; TITLE OF INVENTION: ADJUVANT  
; FILE REFERENCE: N.882328 GCW  
; CURRENT APPLICATION NUMBER: US/10/507,928  
; CURRENT FILING DATE: 2004-09-17  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of the coding insert in p73i-RT3  
US-10-507-928-12

## Alignment Scores:

Pred. No.:	1-56e-13	Length:	562
Score:	301.00	Matches:	145
Percent Similarity:	37.2%	Conservative:	104
Best Local Similarity:	21.6%	Mismatches:	267
Query Match:	2.1%	Indels:	154
DB:	6	Gaps:	26

US-10-723-552-3 (1-8132) x US-10-507-928-12 (1-562)

2628 ATGGGTTGGCAAGCAAGTTCCTCCCAAGTATTCACTGAAGCCAGTGCACACCA 2687  
Db 1 MetGlyProIleSerProIleGluThrValSerValLysLysProGlyMetAspGly 20  
2688 GTGTCAGTCAGACAGTACCCCTTGAGTAAAGAGCTCAAGAAGGAATTCGGCCGCGATGC 2747  
Db ProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCys 40  
2748 CAAGATTAAATCAACAGGGCATCTCTAGTTCCTGTC-----CAATCTCCCTGGAATACT 2801  
Db ThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThr 60  
2802 CCCCTGTACTCCGTTAGAAAGCCTGGGACTAATAGTATCGACAGTACAGGACTTGAGA 2861  
Db ProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 80  
2862 GAGGTCAATAACGGGTGCAGGAT-----ATACACCCCAACAGTCCCGAACCTTAT 2912  
Db GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 100  
2913 AACCTTGTGTGCTCTCCACCCCAACGAGCTGGTATACAGTATTGACTTTAAAGGAT 2972  
Db GlyLeu-----LysLysLysLysSerValThrValLeuAspValGlyAsp 115  
2973 GCCTTCTTCTGCTGAGATTACACCCCACTACCCCACTATTTGCTTCCGATTCGAGAGA 3032  
Db AlatyPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 135



```
QY 3033 GATCCAGGTACGGAGAACCGGG---CAGCTCACCTGGACCCGACCTGCGCCCAAGGGTTC 3089
Db 136 SerIleAsnAenGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 155
QY 3090 AAGAACTCCCGGACCATCTTTGACGAAGCCCTACACAGACACCTGGCCAACTTCAGGATC 3149
Db 156 LysGlySerProAlaIlePheGlnSerMetThrLysIleLeuGluProPheArgLys 175
QY 3150 CAACACCTCAGGTGACCCCTCCAGTACGTGATGACCTGCTCTGCGGAGGACCAAC 3209
Db 176 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrVal---GlySerAsp 194
QY 3210 AAACAGACTGCTTAGAAGCAGCAGAG-----GCATCTACTGCTGGNAATTGCTGACCTA 3263
Db 195 LeuGluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrp 214
QY 3264 GGCTACAGAGCCTCTGCTAAGAGGCCACAGATTTCAGAGGAGAGAGTAACATACTTG--- 3320
Db 215 GlyLeuThrThrProAspLysLysHisGln-----LysGluProProPheLeuTrp 231
QY 3321 ---GGGTACATTTGCGGACGGGACGCGATGCTGACGAGGACGAGCAAGAAACTGTA 3377
Db 232 MetGlyTyrGluLeuHisPro---AspLysTrp-----ThrVal 243
QY 3378 GTCCAGATACGGCCCAACCC-----ACAGCCAAACAAATGAGAGAGTTTGTG 3425
Db 244 GlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuVal 263
QY 3426 GGGACAGCTGGATTTTCAGACGTGTGGATCCCGGGGTTT----- 3464
Db 264 GlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuLysLys 283
QY 3465 -----GCGACCTTAGCAGCCCACTTACCCGCTAACCCAA----- 3500
Db 284 LeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeu 303
QY 3501 -----CAAAAGGGGAAATCTCTGGGCTCTGAGCAGCAGGAGGATTTGATGCT 3551
Db 304 GluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAspPro 323
QY 3552 ATCAAAAGGCCCTGTGAGCGCACCTGTCTGCGCCCTCCCTGACGTAACTAAACCCCTTT 3611
Db 324 SerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTrpThr-----Tyr 341
QY 3612 ACCTTTATGTGATGAGCTAGAGGATAGCCCGGGAGTTTAAACCCAAACCTTAGGA 3671
Db 342 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 361
QY 3672 CCATGGAGAGACCTGTGCGCTACCTGTCAAGAGAGCTCGATCCTGTAGCAGTGGTGG 3731
Db 362 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr----- 379
QY 3732 CCATATGCTGAAGGCTATCGCAGCTGTGGCCATCTGCTCAAGGACGCTGACAAATTG 3791
Db 380 -----GluSerIleValIleTrpGlyLys----- 387
QY 3792 ACTTTGGGACAGATATACTGTAATAGCCCCCATGCAATGCGAGACATCGTTGGCAG 3851
Db 388 -----ThrProLysPheLysLeuProIleGlnLysGlu 398
QY 3852 CCCCCAGCATGATGATGACCAACGCCCGCATGACCCACTATCAAGCCCTGCTTCTCACA 3911
Db 399 ThrTrpGluThrTrpTrpTrpGlu-----TyrTrpGln----- 409
QY 3912 GAGAGGGTCAAGTTGCTGCCACAGCCGCTCTCAACCCCTGCCACTCTCTGCTCGTAA--- 3968
Db 410 -----AlaThrTrpIleProGluTrp 416
QY 3969 -----GAGACTGATGAACCAAGTACTCATGATTCATCACTATTGATGATGAGAGACT 4022
Db 417 GluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGlu----- 434
```

```
QY 4023 GGGGTCCGCAAGGACCTTTACAGACATACCGCTGACTGGAGAACTGCTAACTGGTTCACT 4082
Db 435 -----ProIleValGly-----AlaGluThrPheTyrVal 444
QY 4083 GACGGAAGCAGC-----TATGTGGTGAAGGTAAAG 4112
Db 445 AspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGly 464
QY 4113 AGATGGCTGGGGCGGGTGGTGGACGGACCGCACGATCTGGGCCACGACGCTGCCG 4172
Db 465 ArgGlnLysValValThrLeuThrAsp----- 473
QY 4173 GAAGGAATTCAGCACAAAGGCTGAGCTCATGCCCTCAGCAAGCTTTGCGGCTGGCC 4232
Db 474 -----ThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGln--- 490
QY 4233 GAAGGAAATCCATAAACATTTATACGACAGCAGGATGCTTTCGACATGCACACGTA 4292
Db 491 SerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGly----- 506
QY 4293 CATGGGGCCATCTATAAACAAAGGGGTGCTTACCTCAGCAGGAGGGGAAATAAAGAAC 4352
Db 507 -----IleIleGlnAlaGlnProAspGlnSerGlu 516
QY 4353 AAAGAGGAATTCCTAAGCCTATTAGAACCGCTACATTTACCAAAAAGGCTAGCTATTATA 4412
Db 517 SerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGluLysValTyrLeuAla 536
QY 4413 CACTGTCTGACATCAGAAAGCTAAAGATCTCATATCCAGAGAAACCAAGATGGGTGAC 4472
Db 537 TrpValProAlaHisLysGlyIle-----GlyGlyAsnGluGlnValAsp 551
QY 4473 CGGGTTCAGCAGCAGGAGCCCGAGGGTGT 4502
Db 552 LysLeuValSerAlaGlyIleArgLysVal 561

RESULT 9
US-11-029-465-10
; Sequence 10, Application US/11029465
; Publication No. US20050256070A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Ralph P.
; APPLICANT: Thomsen, Lindy
; APPLICANT: Van-Wely, Catherine
; APPLICANT: Ertl, Peter
; TITLE OF INVENTION: Adjuvant
; FILE REFERENCE: 033267-015
; CURRENT APPLICATION NUMBER: US/11/029,465
; CURRENT FILING DATE: 2005-01-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of RT insert of p7077-RT3
US-11-029-465-10

Alignment Scores:
Pred. No.: 1,56e-13 Length: 562
Score: 301.00 Matches: 145
Percent Similarity: 37.2% Conservativeness: 104
Best Local Similarity: 21.6% Mismatches: 267
Query Match: 2.1% Indels: 154
DB: Gaps: 26

US-10-723-552-3 (1-8132) x US-11-029-465-10 (1-562)
QY 2628 ATGGGTTTGGCAAGCAAGTTCCCCACAAAGTTATTCACTGAAGCCAGTGCACACCA 2687
Db 1 MetGlyProIleSerProIleGluThrValSerValLysLeuLysProGlyMetAspGly 20
```



; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 561  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of the coding insert in  
; OTHER INFORMATION: p731-RT3  
US-11-029-465-12

Alignment Scores:  
Pred. No.: 1.37e-12 Length: 561  
Score: 287.50 Matches: 139  
Percent Similarity: 36.1% Conservative: 105  
Best Local Similarity: 20.6% Mismatches: 265  
Query Match: 2.0% Indels: 167  
DB: 7 Gaps: 24

US-10-723-552-3 (1-8132) x US-11-029-465-12 (1-561)

QY	2628	ATGGGTTTGGCAAGCAAGTTCCCCCAAGTTATTCAACTGAAGCGCAGTGCACACCA	2687
DB	1	MetGlyProIleSerProIleGluThrValSerValLysLeuLysProGlyMetAspGly	20
QY	2688	GTGTCAGTCAGACAGTACCCCTTCAGTAAAGAACTCAAGAAAGAAATTCGGCCGCATGTC	2747
DB	21	ProLysValLysGln---ProLeuThrGluGluLysLysLysAlaLeuValGluLeuCys	39
QY	2748	CAAGATTAAATCCAAAGGGGCATCTAGTTCTCTGTC-----CAATCTCCCTGGAACT	2801
DB	40	ThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThr	59
QY	2802	CCCTGCTACCGTTAGAACCTGGGACTAATGACTATCGACACGATACAGGACTTGAGA	2861
DB	60	ProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg	79
QY	2862	GAGGTCAATAACGGGTGCAGAT-----ATACACCAACAGCTCCCGAACCTTAT	2912
DB	80	GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla	99
QY	2913	AACCTCTGTGTCTCTCCCAACCGGAGCTGGTATACAGTATTGAGCTTAAAGGAT	2972
DB	100	GlyLeu-----LysLysLysLysSerValThrValLeuAspValGlyAsp	114
QY	2973	GCCTTCTTCCTCCGTAGATACACCCCACTAGCAACCACTTTTGGCTTCGATGGAGA	3032
DB	115	AlaTyrPheSerValProLeuAspGluAspPheArgLysTrpThrAlaPheThrIlePro	134
QY	3033	GATCCAGTACGGGAAGAACCGGG---CAGCTCACCTGGACCCGACTGCCCAAGGGTTC	3089
DB	135	SerIleAsnLysGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp	154
QY	3090	AAGAACTCCCGACCATTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATC	3149
DB	155	LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuLeuGluProPheArgLys	174
QY	3150	CAACACCTCAGGTGACCTCTCCAGTACGTGATGACCTGCTCTGGCGGGAGCCACC	3209
DB	175	GlnAsnProAlaIleValIleTyrGlnTyrMetAspLeuTyrValGly-----	191
QY	3210	AAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGTCTGGAATTTGCTGACTAGGCTAC	3269
DB	192	-----SerAspLeuGluIle	196
QY	3270	AGAGCCTCTGCTAAGAGGCCCAAGATTTCAGGAGAGAGGTAAACATCTTGGGGTACAGT	3329
DB	197	GlyGlnHisThrArgLysIleGluGluLeuArgGlnHisLeuLeuAspGlyTrpGlyLeuThr	216
QY	3330	TTGGCGGAGCGGACGCGA-----TGCGTG-----ACGAG	3359
DB	217	ThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHis	236
QY	3360	GCACGGAAGAAAACCTGTAGTCCAGATACCGGCCCAACC-----ACAGCCAAA	3407

DB	237	ProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsn	256
QY	3408	CAAAATGAGAGAGTTTTTGGGGCAGAGCTGGATTTTGGACAGCTGGATCCCGGGGTTT---	3464
DB	257	AspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLys	276
QY	3465	-----GCGACCTTAGCAGCCCTACTACCCGCTA	3494
DB	277	ValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeu	296
QY	3495	ACCAAA-----GAAAAAGGGGAATTCCTCGGCTCTCTGAGCAC	3533
DB	297	ThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis	316
QY	3534	CAGAGGCATTTGATGCTATCAAAAAGGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCT	3593
DB	317	GlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly	336
QY	3594	GACGTAACTAAACCTTTTACCTTTATGTGATGAGCGTAAGGAGTAGCCCGGGGAGTT	3653
DB	337	GlnTrpThr-----TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys	354
QY	3654	TTAACCCCAACCTTAGGACCATCGAGAGACCTGTCTGCTTACCTGTCAAAAGAACTCGAT	3713
DB	355	TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln	374
QY	3714	CTGTAGCCAGTGGTGGCCCATATGCTGAAGCTATCGAGCTGTGGCCATCTGGTC	3773
DB	375	LysIleThrThr-----GlutIleValIleTrpGly	385
QY	3774	AAGGACCTGACAAATTGACTTTTGGGACAGATATATACTGTAAATAGCCCCCATGATTG	3833
DB	386	Lys-----ThrProLysPheLys	391
QY	3834	GAGAACTCCTTCCGACACCCCGACCGATGGATGACCAACGCCCATGACCCACTAT	3893
DB	392	LeuProIleGlnLysGluThrTrpGluThrTrpTrpGlu-----TyrTrp	407
QY	3894	CAAGCCTGCTTCTCACAGAGAGGTCAGTTCGCTCCACCGCGGCTCTCAACCCCTGCC	3953
DB	408	Gln-----Ala	409
QY	3954	ACTCTTCTGCTGAA-----GAGACTGATGAACAGTGCATCATGTTGCCATCAA	4004
DB	410	ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln	429
QY	4005	CTATTGATTGAGGAGACTGGGTCCCGCAGAGACCTTACAGACATACCGCTGACTGGGAA	4064
DB	430	LeuGluLysGlu-----ProIleValGly---	437
QY	4065	GTGCTAACCTGGTTCACCTGACGGAGCAGC-----	4094
DB	438	AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly	457
QY	4095	TATGTGTGAAGGTAAAGAGGATGGCTGGGCGCGGTGGTGACGGACGGACCCGACCATC	4154
DB	458	TyrValThrAsnArgGlyArgGlnLysValValThrLeuThrAsp-----	472
QY	4155	TGGGCCAGCAGCTGCCGGAAGAACTTCAGCAAAAGCGCTGAGCTCATGGCCCTCAGC	4214
DB	473	-----ThrThrAsnGlnLysThrGluLeuGlnAlaIleTyr	484
QY	4215	CAAGCTTTGGGCTGGCCGGAAGGAATCCATAACATTTATACGGACGACGAGTATGCC	4274
DB	485	LeuAlaLeuGln---AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAla	503
QY	4275	TTTCCGACTGCACAGCTACATGGGCCCATCTATAACAAAGGGGGTGTCTTACCTCAGCA	4334
DB	504	LeuGly-----IleIleGlnAla	509
QY	4335	GGGAGGGGAATAAAGAACAAAGAGGAAATTTCTAAGCCTATTAGAGCGGTATATTACCA	4394

Db 510 GlnProAspGlnSerGluSerGluLeuValAsnGlnIleleleGluGlnLeuLeuLysLys 529  
QY 4395 AAAAGGCTAGTATTACATCTCTCGACATCAGAAAGCTAAAGATCTCATATCCAGA 4454  
Db 530 GluLysValTyrLeuAlaTyrValProAlaHisLysGlyLe-----Gly 544  
QY 4455 GGAACACAGATGGCTGACCGGGTGGCCAGCAGCAGCCAGCGGTGT 4502  
Db 545 GlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 560  
RESULT 11  
US-11-014-629-1  
; Sequence 1, Application US/11014629  
; Publication No. US20050244376A1  
; GENERAL INFORMATION:  
; APPLICANT: Hall, Frederick L.  
; TITLE OF INVENTION: TARGETED GENE DELIVERY TO EXTRACELLULAR MATRIX  
; FILE REFERENCE: 30863-704.302  
; CURRENT APPLICATION NUMBER: US/11/014,629  
; CURRENT FILING DATE: 2004-12-15  
; PRIOR APPLICATION NUMBER: US 08/837,223  
; PRIOR FILING DATE: 1997-04-10  
; PRIOR APPLICATION NUMBER: US 09/904,923  
; PRIOR FILING DATE: 2001-07-13  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: moloney murine leukemia virus  
US-11-014-629-1  
Alignment Scores:  
Pred. No.: 4.35e-10 Length: 229  
Score: 250.00 Matches: 70  
Percent Similarity: 39.1% Conservative: 23  
Best Local Similarity: 29.4% Mismatches: 75  
Query Match: 1.7% Indels: 70  
DB: 1.7% Gaps: 12

US-10-723-552-3 (1-8132) x US-11-014-629-1 (1-229)  
QY 5776 CATAAACCCCTATCTCTACCTGGTTAAATTAAGTACTCGGCACAGTAAATATATCAAC 5835  
Db 8 HisGlnValTyrAsnIleThrTrpGluValThr---AsnGlyAspArgGluThrValTrp 26  
QY 5836 AACACTCAAGGAGGCTCTTTAGGAACCTGTGGCTGATCTATAC-----GTTTGC 5889  
Db 27 AlaThrSerGlyAsnHisProLeuTrpThrTrpProAspLeuThrProAspLeuCys 46  
QY 5890 CTC-----AGATCAGTTATTCCTAGT 5910  
Db 47 MetLeuAlaHisHisGlyProSerTyrTrpGlyLeuGluTyrGlnSerProPheSerSer 66  
QY 5911 CTGACCTCACCCCA----- 5925  
Db 67 ProProGlyProProCysCysSerGlySerProGlyCysSerArgAspCysGlu 86  
QY 5925 ----- 5925  
Db 87 GluProLeuThrSerLeuThrProArgCysAsnThrAlaTrpAsnArgLeuLysLeuAsp 106  
QY 5926 GATATCTCCATGCT-----CACGGATTTTANGTTTGGCCAGGACCA-----CCAAAT 5973  
Db 107 GlnThrThrHisLysSerAsnGluGlyPheTyrValCysProGlyProHisArgProArg 126  
QY 5974 AATGGAACCATTCGCGAAATCCAGAGATTTCTTTTGTAAACAATGGAAGTGTGAACC 6033  
Db 127 GluSerLysSerCysGlyGlyProAspSerPheTyrCysAlaTyrTrpGlyCysGluThr 146  
QY 6034 TCTAATGATGATATTGGAAATGGCAACCTCTCAGCAGATAGGTAAGTTTCTTAT 6093  
Db 146 TCTAATGATGATATTGGAAATGGCAACCTCTCAGCAGATAGGTAAGTTTCTTAT 6093

Db 147 ThrGlyArgAlaTyrTrpLys---ProSerSerSerTrpAsp-----PheIleThr 162  
QY 6094 GTCAACACACCTATACCACTCTGACCAATTTAATTACCTGACCTGGATTAGCACTGGAGC 6153  
Db 163 ValAsnAsnAsnLeuThrSerAspGln-----AlaValGlnValCysLysAspAsn 179  
QY 6154 CCAAGTGTCTCTCTTCAGACCTAGATTACCTAAATAAGTTTCACTGAGAAGGAAAA 6213  
Db 180 LysTrpCysAsnPro-----LeuValIleArgPheThrAspAlaGlyArg 194  
QY 6214 CAAGAAATATCTCTAAATGGTAAATGGTATCTCTGGGAATGGTATATTATGGAGGC 6273  
Db 195 Arg-----ValThrSerTrpThrThrGlyHisTyrTrpGlyLeuArgLeuTyr---Val 211  
QY 6274 TCGGTAAACCAACCAAGCTCCATTTCTAACTATTTCGCTCAAAATAAACCAAGCTG 6327  
Db 212 SerGlyGlnAspProGlyLeuThrPheGlyIleArgLeuArgTyrGlnAsnLeu 229  
RESULT 12  
US-11-022-562-221  
; Sequence 221, Application US/11022562  
; Publication No. US20050249742A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruprecht, Ruth M.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING  
; FILE REFERENCE: DFN-043CN  
; CURRENT APPLICATION NUMBER: US/11/022,562  
; CURRENT FILING DATE: 2004-12-22  
; PRIOR APPLICATION NUMBER: PCT/US03/20322  
; PRIOR FILING DATE: 2003-06-27  
; PRIOR APPLICATION NUMBER: 60/392718  
; PRIOR FILING DATE: 2002-06-27  
; NUMBER OF SEQ ID NOS: 340  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 221  
; LENGTH: 210  
; TYPE: PRT  
; ORGANISM: Human Immunodeficiency Virus  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 31, 97, 140, 141, 144, 178  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-11-022-562-221  
Alignment Scores:  
Pred. No.: 7.91e-08 Length: 210  
Score: 217.50 Matches: 65  
Percent Similarity: 47.9% Conservative: 37  
Best Local Similarity: 30.5% Mismatches: 82  
Query Match: 1.5% Indels: 29  
DB: 1.5% Gaps: 9

US-10-723-552-3 (1-8132) x US-11-022-562-221 (1-210)  
QY 2784 CAATCTCCCTGGAATCTCCCTGCTACCGTTAGAAACCTGGACTAATGACTATCGA 2843  
Db 16 GluAsnProTyrAsnThrProValPheAlaIleLysLysAsp\*\*\*ThrLysTrpArg 35  
QY 2844 CCAGTACAGAGCTTGAGAGAGGTCAATAAACGGGTGCAGGAT-----ATACACCCA 2894  
Db 36 LysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeu 55  
QY 2895 ACAGTCCGGAACCTTATAACCTTGTGTCTCTCCCAACCCCAACGAGCTGGTATACA 2954  
Db 56 GlyIleProHisProAlaGlyLeu-----LysLysLysLysSerValThr 70  
QY 2955 GTATTGGACTTAAAGATGCTCTTCTGCTGAGATTACCCCACTAGCAACCACTT 3014  
Db 71 ValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyr 90  
QY 3015 TTTGCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGG---CAGCTCACCTGGACC 3071

```

Db      91  ThrAlaPheThrIlePro***ThrAsnAsnGluThrProGlyIleArgTyrGlnTyrAsn 110      :
Qy      3072 CGACTGCCCAAGGTTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAC 3131      :
Db      111  ValLeuProGlnGlyTyrLysGlySerProAlaIlePheGlnSerSerMetThrLysIle 130      :
Qy      3132 CTGCCCAACTTCAGATCCAAACCCCTCAGTGACCTCTCCAGTACGTGGATGACCTG 3191      :
Db      131  LeuGluProPheArgIleLysAsnPro*****ValIle***GlnTyrMetAspAspLeu 150      :
Qy      3192 -----CTTCTGCGCGGAGCCACCAACAGGACTCTTAGAAGGCGACCAAG 3236      :
Db      151  TyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgLys 170      :
Qy      3237 GCATCTACTGTGGAAATGTCTGACTAGGTACAGACCTCTGTCTAAGAGGCCACAGATT 3296      :
Db      171  HisLeuLeu-----SerTrpGlyPhe***ThrProAspLysLysHisGln--- 185      :
Qy      3297 TGCAGGAGAGAGGTAAACATACTTG-----GGGTACAGTTTGGCGGACGGCGACGATGG 3350      :
Db      186 -----LysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro---AspLysTrp 202      :
Qy      3351 CTGACGAGGCGACCGAAGAAACTGTAGTCCAGATACCG 3389      :
Db      203  ThrValGlnPro-----IleGlnLeuPro 210

RESULT 13
US-11-230-251-26
; Sequence 26, Application US/11230251
; Publication No. US20060019322A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Hervé
; APPLICANT: Cafferey, Robert
; APPLICANT: Ali, Shujath
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific
; TITLE OF INVENTION: Genes
; FILE REFERENCE: DEX-0239
; CURRENT APPLICATION NUMBER: US/11/230,251
; CURRENT FILING DATE: 2005-09-19
; PRIOR APPLICATION NUMBER: US/09/957,708
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/233,746
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-230-251-26

Alignment Scores:
Pred. No.:      3,03e-07      Length:      422
Score:          210.50      Matches:      88
Percent Similarity: 39.3%      Conservative: 52
Best Local Similarity: 24.7%      Mismatches: 109
Query Match:      1.4%      Indels:      107
DB:              7          Gaps:      18

US-10-723-552-3 (1-8132) x US-11-230-251-26 (1-422)
Qy      4188 CAAAGGCTAGCTCAGTGGCCCTCAGCAAGCTTTGGGTGGCGGAGGGAATCCATA 4247
Db      14  GlnArgAlaGluLeuValAlaValIleThrValLeuGln---AspPheAsnGlnSerIle 32
Qy      4248 AACATTTATCCGACACAGGTATGCCTTT---CGGACTGCACACGTACATGGGGCCATC 4304
Db      33  AsnIleValSerAspAlaTyrValValGlnAlaThrLysAspIleGluArgAlaLeu 52
Qy      4305 TATAACAAAGGGGGTGTCTTACCTCAGCAGGAGGGAATATAAGAACAAAGAG----- 4358

```

```

Db      53  IleIysTyr-----IleMetAspAspGlnLeuAsn 62
Qy      4359 GAAATTTCTAAGCCTATTAGAGCCGTACATTTACCAAAAAGG-----CTAGCTATT 4409
Db      63  ProLeuPheAsnLeuGlnGln---AsnValArgLysArgAsnPheProPheTyrIle 81
Qy      4410 ATACACTGTCTCGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCCAGATGGCT 4469
Db      82  ThrHisIleArgAlaHisThrAsnLeuProGlyProLeuThrArgAlaAsnGluGlnAla 101
Qy      4470 GACCGGGTTTCCCAAGCAGGCGCCAGGGTGTTAACCTTCTGCCTATAATAGAAAATGCC 4529
Db      102  AspLeuLeuValSerSerAla----- 108
Qy      4530 AAAGCCCCAGAACCCAGACAGTACACCTAGAGACTGGCCAGAGAGATAAAAAAGATA 4589
Db      108 ----- 108
Qy      4590 GACCAGTTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATCCTG 4649
Db      109 -----PheMetGlu----- 111
Qy      4650 CCCACAAAGAAGGGTTAGAATATGTCACACAGATACATCGTCTAACCCAC----- 4700
Db      112 -----AlaGlnGluLeuHisAlaLeuThrHisValAsnAla 123
Qy      4701 CTAGGAACCTAAACACACCTGCAGCAGTGGTCCAGAACATCCCTTATCATGTCTGAGCTA 4760
Db      124  IleGlyLeuLysAsnLysPheAspIleThrTrpLysGln----- 136
Qy      4761 CCAGGAGTGGCTGACTCGGTGGTCAAAACATTTGTGTCCTCCAGCTGGTGTAAAT----- 4814
Db      137 -----ThrLysAsnIleValGlnHisCysThrGlnCysGlnIleLeuHisLeuAla 153
Qy      4815 GCTAATCTTCCAGAATGCCCTCCAGGAAGAGACTAAGGGGAAGCCACCCAGCGCTCAC 4874
Db      154  ThrGlnGluAlaArgValAsnPro-----ArgGlyLeuCysProAsnValLeu 169
Qy      4875 TGGGAAGTGCATCTTACTAGGTAAAGCCGGCTAAATACCGAAACAATACCTATTGTT 4934
Db      170  TrpGlnMetAspValMetHisVal---ProSerPheGlyLysLeuSerPheValHisVal 188
Qy      4935 TTTGTAGACACCTTTTCAGGATGGTAGAGGCT---TATCCTACTAAGAAAGAGACTTCA 4991
Db      189  ThrValAspThrTyrSerHisPheIleTrpAlaThrCysGlnThrGlyGluSerThrSer 208
Qy      4992 ACCGTGTGGCTAAAAAATACTCGAAGAAATTTTCCAGATTTTGGAAATCTCAAGGTA 5051
Db      209  HisVal-----LysArgHisLeuLeuSerCysPheProValMetGlyValProGluLys 226
Qy      5052 ATAGGGTCACACAATGGTCCAGCTTTTGTGGCCAGGTAAAGTCCAGGACTGGCCCAAGATA 5111
Db      227  ValLysThrAspAsnGlyProGlyTyr-----CysSerLysAlaValGlnLysPhe 243
Qy      5112 TTGGGGATTGATTGGAAA-----CTGCATTGTGCATACACACCCCAAGCTCA 5159
Db      244  LeuAsn---GlnTrpLysIleThrHisThrIleGlyIleLeuTyrAsnSerGlnGlyGln 262
Qy      5160 GGACAGGTAGAGAGATGAATAGAACCATTAAGAGACCTTACTATAA 5207
Db      263  AlaIleIleGluArgThrAsnArgThrLeuLysAlaGlnLeuValLys 278

RESULT 14
US-11-124-367A-444
; Sequence 444, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CU001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A

```







```
QY 4485 CAGCAGCCAGGGTGTAACTTCTGCTATATAGAAATGCCAAAGCCCCAGAACCC 4544
|||
1445 Gln-----LysAspLysIleProGlu--- 1451
|||
QY 4545 AGACGACGTACACCTAGAACACTGSCAAGAGATAAAGATAGACCACTTCTCTGAG 4604
|||
1452 -----GluLysAspLysAlaLeuGluGln----- 1459
|||
QY 4605 ACTCCGAAGGACCTGCTATACCTCAGATGGGAAGAAATCTCTGCCCAACAAAGAGG 4664
|||
1460 -----LysAspThrAlaLeuGluGlnLysAspLysAlaLeuGluProLysAspLysAsp 1477
|||
QY 4665 TTAGAA-----TATGTCCACAGATACATGCTCTAAACCCACCTAGGAACATAA 4712
|||
1478 LeuGluGlnLysAspArgValLeuGluGlnLysGluLysIleProGluGluLysAspLys 1497
|||
QY 4713 CACTGACGAGTGTGTCAGAACATCCCTTATCATGTTCTGAGGCTACAGAGTGGCT 4772
|||
1498 AlaLeuAspGlnLysValArgSer----- 1505
|||
QY 4773 GACTCGGTGTCAACATGTGTGCTCTGCCCTGCGAGCTGTAACTTAATCTTCCAGA--- 4829
|||
1506 -----ValGluHisLysAlaProGluAspThrValAlaGluMetLysAspArgAsp 1522
|||
QY 4830 ---ATGCTCCAGGAAGAGACTAAGGGGAAGCCACCGCGCTCACTGGGAAGTGGAC 4886
|||
1523 LeuGluGlnThrAspLysAlaProGluGlnLysHisGlnAlaGlnGluGlnLysAspLys 1542
|||
QY 4887 TTCCTGAGTAAAGCCGGCTAAATACGGGAACAAATACCTATTTGGTT----- 4934
|||
1543 ValSerGluLysLysAspGlnAlaLeuGluGlnLysTyrTrpAlaLeuGluGlnLysAsp 1562
|||
QY 4935 -----TTTGTAGACACCTTTTCCAGGATGGGTAGAGCTTATCTCTACTAAGAAGAGACT 4988
|||
1563 GluAlaLeuGlnGlnAlaLeuGluGlnLysHisGlnHisGlnThrGlnGlnGluGln 1582
|||
QY 4989 TCAACCGTGTGGCTAAAAAATACCTGGAAGAAATTTTCCAAAGATTGGAATACCTAAG 5048
|||
1583 SerLeuValGlnGluAspLysThrArgLys-----ProLys 1594
|||
QY 5049 GTAATAGGTGCACAAATGTCTCAGCTTTTGTGTCGCCAGTAAAGTACGAGACTG--- 5102
|||
1595 MetLeu---GluGluLysSerProGluLysValLysAlaMetGluGluLysLeuGluAla 1613
|||
QY 5103 -----GCCAAGATATTGGGATTTGGAATGGAACCTGCATTGTGCATACAGACC 5150
|||
1614 LeuLeuGluLysThrLysAlaLeuGluLysLeuGluGlnSerLeu-----ValGlnGluGly 1631
|||
QY 5151 CAAAGCTCAGCAGCAGGTAGAGAGATGAATAGAACCAATTAAGAGAGACCTTACTAAATTG 5210
|||
1632 ArgAlaArgGluGlnGluLysTyrTrpArg----- 1642
|||
QY 5211 ACCCGGAGACTCGGCTTAATGATTTGATGAGTCTCTGCTGCTGCTGCTTTTGGT 5270
|||
1643 ---GlyGlnAspValValGlnGluLysTrpGlnLysSerPro-----Thr 1656
|||
QY 5271 AGGAACACCCCTGACAGATTGGGCTGACCCCTATGAAATTAATCTACGGGGACCCCCC 5330
|||
1657 ArgGluGluProAlaGly-----GluGlnLysGluLeu-----Ala 1668
|||
QY 5331 CCATTTGTAATAATGCTTCTGTACATAGT-----GCTGAC 5366
|||
1669 ProAlaTrpGluAspThrSerProGluGlnAspAsnArgTyrTrpArgGlyArgGluAsp 1688
|||
QY 5367 GTGCTGCTTCCAGGCTTTGCTCTAGGCTCAAGGCACTTGAAG-----TGGGTG 5417
|||
1689 ValAlaLeuGluGlnAspThrTyrTrpArgGluLeuSerCysGlnArgLysValTrpPhe 1708
|||
QY 5418 AGACACGAGCGTGGAGGCACTCCGGAGGCT-----ACTCAGGAGGAGACT-- 5469
|||
1709 -ProHisGluLeuAspGlyGlnGlyAlaArgProHisTyrThrGluGluArgGluSerTh 1728
|||
```

```
Db      1982 aCysProThrArgGluPro-ProLeuGlyAlaAlaGlyAspTrpProCysLeuSerT 2002
      ||| ||||| ||| |||
Qy      6441 CAGCAGCACTAAATGGGGCAAACTTTTGTAGCTCATCCAGGAGCTTTTCAAGCTCT 6500
      ||| ||||| ||| |||
Db      2002 hrlYsGluAlaAlaGlyArgAsn---ThrSerAlaGluLysGluLeuSerProI 2021
      ||| ||||| ||| |||
Qy      6501 TAACTCCAGCA---CTCAGAGGCTACTCTTCTTGTGTGGCTATGCTTAGCTTCGGGCCC 6557
      :|||: |||: |||: |||:
Db      2021 leSerProLysSerLeuGlnSerAspThrProThrPheSerTyrAlaAlaLeu---AlaG 2040
      ||| ||||| ||| |||
Qy      6558 ACCTTACTATCAGGAATGGCTAGACAGAGGGAATCAATGTGTGCAAAAGAACATAGAGA 6617
      ||| ||||| ||| |||
Db      2040 lyProThrValProArgProGluProGlyProSerMet----- 2053
      ||| ||||| ||| |||
Qy      6618 CCAATGCACATGGGATCCCAAAATAAGCTTTACCTTTACTGAGGTCTTCTGGAAGGCAC 6677
      :|||: |||: |||: |||:
Db      2054 -----GluProSerLeuThrProAlaValPro-----ProArgAlaP 2067
      ||| ||||| ||| |||
Qy      6678 CTGCATAGGAAGGTTCCCATCCACACACCTTTGTAAACCACACTGAAGCCTTTAA 6737
      ||| ||||| ||| |||
Db      2067 rotleLeuSerLysGlyProSerPro----- 2075
      ||| ||||| ||| |||
Qy      6738 TCAAACTCTGAGAGTCAATATC 6760
      ||| ||||| ||| |||
Db      2076 -----ProLeuAsnGlnYAsnIle 2081

RESULT 15
US-11-124-367A-442
; Sequence 442, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; FILE REFERENCE: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 442
; LENGTH: 2803
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-367A-442

Alignment Scores:
Pred. No.:      3,18e-06      Length:      2803
Score:          199.50      Matches:      425
Percent Similarity: 31.1%      Conservative: 257
Best Local Similarity: 19.4%      Mismatches: 787
Query Match:      1.4%      Indels:      722
DB:              7          Gaps:      105

US-10-723-552-3 (1-8132) x US-11-124-367A-442 (1-2803)
Qy      1296 CTCACGGGGTGGTGGAGTCCCTTATGTTCTCCACGAGCTACTTGGGATGTTGCAA 1355
      :||| ||||| ||| |||
Db      247 ileThrAlaLeuVal-----ValtrpLeuProAlaAsnProThr-----Glu 260
      ||| ||||| ||| |||
Qy      1356 CAGCTGCTGCAGACACTCTTTCACAAACGAGGAGCGAGAGAGAATTCATTAGAGGCTAGA 1415
      :|||: |||: |||: |||:
Db      261 lysileValArgValLeuPheProGlyAsnAlaProGlnAsnLysIleLeuGluGlyLeu 280
      :|||: |||: |||: |||:
Qy      1416 AAAAATGTTCTCTGGGCGGCAC-----GGGCGCACCCAGCGGTTTGC AAAATGAGATTGAC 1469
      :|||: |||: |||: |||:
```

```
Db      281 GluLysLeuArgHisLeuAspPheLeuArgTyrProValAlaThrGlnLysAspLeuAla 300
Qy      1470 ATGGGATTTCCCTTAAT-----CGCCCCGGTTGGAGCTACAAACGGCTGAAGGTAGG 1523
      ||| ||||| ||| |||
Db      301 SerGlyAlaValProThrAsnLeuLysProSerLysIleLysGlnArgAlaAspSerLys 320
      ||| ||||| ||| |||
Qy      1524 GAGAGCTTTGAAATCTATCGCCAGGCTCTGCTGGCGGCTCTC----- 1565
      ||| ||||| ||| |||
Db      321 GluSerLeuLysAlaThrThrLysThrAlaValSerLysLeuAlaLysArgGluGluVal 340
Qy      1566 -----CGGGCGCCTCAAGACGGCCCACTAAATTTGGCTAAG-----GTA 1604
      ||| ||||| ||| |||
Db      341 ValGluGluGlyAlaLysGluAlaArgSerGluLeuAlaLysGluLeuAlaLysThrGlu 360
Qy      1605 AGAGAGTGTATGAGGACCGAATGAACCCCTCTCTGTTTTCTTGAGAGGCTCTTGAA 1664
      :|||: |||: |||: |||:
Db      361 LysLysAlaLysGluSerSerGluLysProPro-----GluLysProAlaLys 376
      ||| ||||| ||| |||
Qy      1665 GCCTTCAGGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAGCCTCAGTGGCT 1724
      ||| ||||| ||| |||
Db      377 ProGluArgValLysThrGluSerSerGluAlaLeuLysAlaGluLysArgLysLeuIle 396
Qy      1725 TTGGCCTTTATAGACACAGCTCAGCTTGGATATTAGAAAGCTTCAGAGACTGGAAGGG 1784
      :|||: |||: |||: |||:
Db      397 LysAspLysValGlyLysLysHisLeu-----LysGluLysIleSerLysLeuGlu 414
      ||| ||||| ||| |||
Qy      1785 TTACAGGAGGCTGAGTTACGTATCTAGTCAAGGAGCGCAGAAAGTATATTACAAAGG 1844
      :|||: |||: |||: |||:
Db      415 LysLysAspLysGluLysLysGluIleLysLysGluArgLysGluLeu-----Lys 431
Qy      1845 GAGACAGAAGAAGAGGGAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1904
      :|||: |||: |||: |||:
Db      432 LysAspGluLysArgLysGluLysLysAspAlaLysLysGluLysLys-----ArgLys 450
Qy      1905 CGTAATAACCGCAAGAGAGAAATTTGACTAAG----- 1937
      ||| ||||| ||| |||
Db      451 AspThrLysProGluLeuLysLysIleSerLysProAspLeuLysProPheThrProGlu 470
Qy      1938 -----ATCTTGGCTGCAGTGGTTGAAGGGAAAGCAATACGGAAGAGAGAGA 1985
      :|||: |||: |||: |||:
Db      471 ValArgLysThrLeuTyrLysAlaLysValProGlyArgValLysIleAspArgSerArg 490
      ||| ||||| ||| |||
Qy      1986 GATTTTAGG-----AAAATAGGTGAGCCCTAGACACTCAGGG 2024
      :|||: |||: |||: |||:
Db      491 AlaIleArgGlyGluLysGluLeuSerSerGluProGlnThrProProAlaGlnLysGly 510
Qy      2025 AACTCTG-----GGCAATAGG-----ACCCCTACTCGAC 2051
      :|||: |||: |||: |||:
Db      511 ThrValProLeuProThrIleSerGlyHisArgGluLeuValLeuSerSerProGluAsp 530
Qy      2052 AAGGACCAATGTGCATATTGTAAGAAAGAGGACACTGGGCAAGGAAGTGCCTCCCAAGAG 2111
      ||| ||||| ||| |||
Db      531 LeuThrGlnAspPheGluGluMetLysArgGluGluArgAla----- 544
      ||| ||||| ||| |||
Qy      2112 GGAACAAAGGACCAAGGATCTAGCTCTAGAAAGATATAAGATTAGGGAGACGGGT 2171
      ||| ||||| ||| |||
Db      545 -----LeuLeuAlaGluGlnArgAspThrGly---LeuGly 555
Qy      2172 TCGGACCCCTCCCGAGCCCGGTAACCTTTGAAGGTGGAGGGGCAACACGATTGAGTTC 2231
      ||| ||||| ||| |||
Db      556 AspLysProPhePro-----LeuAspThrAlaGluGlyProProSerThrAla 572
Qy      2232 CTGTTGTTATACCGGAGCGAAACATTCACTGCTACTACAGCCATTAGGAAATATAAGAT 2291
      :|||: |||: |||: |||:
Db      573 IledGlnGlyThrProProSerValProGlyLeuGlnGluGluHisValMetLysGlu 592
Qy      2292 AAAAAATCCTGGGTGATGGTCCACAGGGCAACACAGATATCCATGAGTATACCCGAAGA 2351
      ||| ||||| ||| |||
Db      593 LysGlu----- 594
Qy      2352 ACAGTTGACTTGGGAGTGGGACCGGTAACCCACTCTGTTTCTGGTGCATACCTGAGTGCCA 2411
      :|||: |||: |||: |||:
Db      595 -----LeuValProGluValPro 600
```





QY 5904 TCCTAGTCTGACCTCACCCCCAGATATCCTCCATGCTCAGGATTTTATGTTTCCCCAGG 5963  
|||::: |||||  
Db 1853 oProAlaProLeuSerPro-----AlaProGl 1862  
|||  
QY 5964 ACCACCAATATGGAATAATCGCGGAATCCAGAGATTTCTTTTGTAAACAATGGAA 6023  
|||||  
Db 1862 yProProThrProAlaProGluSerHisThrProAlaProPhe----- 1876  
|||  
QY 6024 CTGTGTAACTCTAATGATGATATTGGAATGGCCAACTCTCAGCAGGATAGGGTAAG 6083  
||| |||::: |||  
Db 1877 -----SerTipGlyThrAlaGluTyraSpSerVal-- 1886  
||| |||  
QY 6084 TTTTCTTATGTCACACCTATACCAGCTCTGGACAATTTAATTACCTGACCTGGATTAG 6143  
::: |||  
Db 1887 -----ValAlaAlaValGl 1891  
|||  
QY 6144 AACTGGAAGCCCAAGTGC-----TCTCCTTCAGACCTAGATTACCTAAA 6188  
|||::: |||  
Db 1891 nGluGlyAlaAlaGluLeuGluGlyProTyraSerProLeuGlyLysAspTyraArgLy 1911  
||| |||  
QY 6189 AATAAGTTTCACTGAGAAGAAACAAGAAATATCTTAAATGGGTAAATGGTATGTC 6248  
||| |||::: |||  
Db 1911 s-----AlaGluGlyGluArgGluGlu-----GluGlyArgAl 1922  
|||  
QY 6249 TTGGGGAAATGGTATATTATGAGGCTCGGGTAAACAACCA---GGCTCCATTCTAACTAT 6305  
: ||| |||  
Db 1922 aGluAlaProAspLysSerHisSerLysValProGluAlaSerLysHisAl 1942  
||| |||  
QY 6306 TCGCCTCAAAATAAACCCAGCTGGAGCCTCCAAATGGCT----- 6342  
::: ||| |||  
Db 1942 aThrThrGluProGluInThrGluProGluInArgGluProThrProAspGl 1962  
|||  
QY 6343 -----ATAGGACCAATACGCTCTTGACGGGTCAAGACCCCC 6380  
||| |||  
Db 1962 uArgSerPheGlnTyraAlaSpIleTyraGluInMetMetIleuThrGlyLeuGlyProAl 1982  
||| |||  
QY 6381 AACCCAGGACAGGACCATCTCTTACATACTTCTGGATCAGACCCCACTGAGTCTAA 6440  
||| |||  
Db 1982 aCysProThrArgGluPro-ProLeuGlyAlaAlaGlyAspTyrProProCysLeuSerT 2002  
||| |||  
QY 6441 CAGCAGCACTAAATGGGGCAAACTTTTACCTCATCCAGGAGCTTTTCAAGCTCT 6500  
||| |||::: |||  
Db 2002 hrLysGluAlaAlaAlaGlyArgAsn---ThrSerAlaGluLysGluLeuSerSerProI 2021  
||| |||  
QY 6501 TAACTCCAGCA---CTCCAGAGGCTACCTCTTCTTGTGGCTATGCTTAGCTTCGGGCC 6557  
::: ||| |||  
Db 2021 leSerProLysSerLeuGlnSerAspThrProThrPheSerTyraAlaLeu---AlaG 2040  
||| |||  
QY 6558 ACCTTACTATGAAGGAATGGCTAGAAGAGGAAATTCAATGTGACAAAGAACATAGAGA 6617  
||| |||  
Db 2040 lyProThrValProProArgProGluProGlyProSerMet----- 2053  
|||  
QY 6618 CCAATGCACATGGGATCCAAAATAAGCTTACCTTACGTAGGTTTCTGGAAAAGGCAC 6677  
::: ||| |||  
Db 2054 -----GluProSerLeuThrProAlaValPro-----ProArgAlap 2067  
||| |||  
QY 6678 CTGCATAGGAAGGTTCCCCATCCCAACACACCTTTGTAAACCACACTGAAGCCTTTAA 6737  
||| |||  
Db 2067 rolleLeuSerLysGlyProSerPro----- 2075  
|||  
QY 6738 TCAACCTCTGAGAGTCAATATC 6760  
||| |||  
Db 2076 -----ProLeuAsnGlyAsnIle 2081  
|||

Search completed: February 14, 2006, 17:20:06  
Job time : 220.655 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 14:51:21 ; Search time 9.17949 Seconds  
(without alignments)  
3295.451 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_585\_2156  
Perfect score: 2848  
Sequence: 1 ATGGGACAGCGGTGACGAC.....CTCTAGAGAAGATAAAGAT 1572

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSSWEB spool/US10723552/runat 14022006 125143 12937/app query.fasta\_1  
-DB=PIR -QFMT=fastan -SUPFI=rpr -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MOD=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US10723552 @CGN 1 1 107 @runat 14022006 125143 12937 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPO=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 80:\*  
1: Pirl.\*  
2: Pirl2.\*  
3: Pirl3.\*  
4: Pirl4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1677.5	58.9	520	1 FOLJGL	gag polyprotein -
2	1637.5	57.5	512	1 FOMVGS	gag polyprotein -
3	1477.5	51.9	537	1 FOMVW7	gag polyprotein -
4	1383.5	48.6	538	2 S35474	gag polyprotein -
5	1381.5	48.5	538	2 S70394	gag polyprotein -
6	1379.5	48.4	538	1 FOMVM	gag polyprotein -
7	1377.5	48.4	1784	2 T10532	gag-pol polyprotei
8	1376.5	48.3	538	1 FOMVIM	gag polyprotein -
9	1371	48.1	537	1 FOMVGV	gag polyprotein -
10	1364	47.9	537	1 FOMVRV	gag polyprotein -
11	1359.5	47.7	540	1 FOMVHL	gag polyprotein -
12	1357	47.6	537	1 FOMVMB	gag polyprotein -
13	1347.5	47.3	536	1 FOMVNE	gag polyprotein -
14	1326	46.6	529	1 FOMVDU	gag polyprotein -

15	1312	46.1	532	1 FOMVVB	gag polyprotein -
16	1239	43.5	536	1 FOMVND	gag polyprotein -
17	1238	43.5	468	1 FOMVNU	gag polyprotein -
18	1049	36.8	427	2 A53798	58k membrane-assoc
19	944	33.1	425	1 FOMVGC	gag polyprotein -
20	930.5	32.7	790	1 FOMVHZ	gag-kit polyprotei
21	778	27.3	371	1 FOMVCS	gag polyprotein -
22	727.5	25.5	608	2 A46312	gag polyprotein -
23	720.5	25.3	310	1 FOMVFB	gag polyprotein -
24	606.5	21.3	559	2 A46311	pol polyprotein -
25	573	20.1	763	1 A40831	gag-akt polyprotei
26	514.5	18.1	313	1 FOVDA	gag polyprotein -
27	497	17.5	981	1 FOMVGM	gag-abl polyprotei
28	428.5	15.0	187	1 FOMVSS	gag polyprotein -
29	384.5	13.5	663	1 TWMVRR	protein-tyrosine k
30	328.5	11.5	145	2 A32325	gag polyprotein -
31	194.5	6.8	697	2 A26132	gag-abl-pol polypr
32	170.5	6.0	1638	2 A42091	transcription acti
33	165	5.8	1006	2 T42731	atrophin-1 related
34	164	5.8	2526	2 T20531	hypothetical prote
35	164	5.8	2722	2 T20532	hypothetical prote
36	164	5.8	2738	2 E88320	protein F07Al1.6 [
37	161	5.7	1184	2 G01763	atrophin-1 - human
38	160.5	5.6	551	2 S57447	HPBRII-7 protein -
39	157	5.5	1280	2 T00365	hypothetical prote
40	154	5.4	618	2 T42664	hypothetical prote
41	154	5.4	707	2 A46302	PTB-associated spl
42	150.5	5.3	1263	2 T15496	hypothetical prote
43	149.5	5.2	887	2 G88484	protein F23Fl2.8 [
44	149	5.2	701	2 F48613	gag polyprotein -
45	149	5.2	701	2 D48613	gag polyprotein -

ALIGNMENTS

RESULT 1

FOLJGL

gag polyprotein - gibbon ape leukemia virus  
N;Contains: core protein p10; core protein p12; core protein p15; core protein p30  
C;Species: gibbon ape leukemia virus  
C;Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 09-Jul-2004  
C;Accession: A32595

R;Delassus, S.; Sonigo, P.; Wain-Hobson, S.

Virology 173, 205-213, 1989

A;Title: Genetic organization of gibbon ape leukemia virus.

A;Reference number: A32595; MUID:90051069; PMID:2683360

A;Accession: A32595

A;Molecule type: genomic RNA

A;Residues: 1-520 <DEL>

A;Cross-references: UNIPROT:P21416; UNIPARC:UPI000012AFF; GB:M26927; NID:g332610; PIDN

C;Genetics:

A;Gene: gag

C;Superfamily: mammalian retrovirus gag polyprotein I

C;Keywords: core protein

F;1-125/Product: core protein p15 #status predicted <P15>

F;126-195/Product: core protein p12 #status predicted <P12>

F;196-454/Product: core protein p30 #status predicted <P30>

F;455-520/Product: core protein p10 #status predicted <P10>

Alignment Scores:

Pred. No.: 2.91e-107 Length: 520  
Score: 1677.50 Matches: 330  
Percent Similarity: 71.1% Conservative: 63  
Best Local Similarity: 59.7% Mismatches: 93  
Query Match: 58.9% Indels: 67  
DB: 1 Gaps: 10

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x FOLJGL (1-520)

Qy 1 ATGGGACAGCGGTGACGACCCCTCTAGTTGACTCTCGACCTGCGACTGAAGTTAAA 60

Db 1 MetGlycInAspAsnSerThrProIleSerLeuThrLeuAsnHisTrpArgAspValArg 20





```

QY 61 TCCAGGGCTCATAATTTCTCAGTTCAGTTAAGAGGACCTTGGCAGACTTTCGTGTC 120
Db 21 ThrArgAlaHisAsnLeuSerValLysIleArgLysGluLysTrpGlnThrPheCysSer 40

QY 121 TCTGAATGGCCGACATTCGATGTCGATGGCCATCAGAGGGGACCTTTAAATCTCAGATT 180
Db 41 SerGluTrpProThrPheGlyValGlyTrpProGluGlyThrPheAsnLeuSerVal 60

QY 181 ATCTGGCTGTATAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCGATCAGGAG 240
Db 61 IlePheAlaValLysArgIleValPheGlnGlu---ThrGlyGlyHisProAspGlnVal 79

QY 241 CCCTATATCTTACGTGCGCAGATTTGCGAGAGATCTCTCGCCATCGGTTAAACCATGG 300
Db 80 ProTyrIleValValTrpGlnAspLeuAlaGlnSerProProTrpValProProSer 99

QY 301 CTGAATAGCCCAAGAAAGCCAGGTCGCCGAATTCCTGGCTCTTGGAGAGAAAAACAACAC 360
Db 100 AlalysIleAlaValValSerSerProGluAsnThrGlnGlyProSerAlaGlyArgPro 119

QY 361 TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTACCCCGAG----- 402
Db 120 SerAla-----ProProArgProProIleTyrProAlaThrAspAspLeuLeuLeu 136

QY 403 ATTGAGGAGCCACCGGCTTGGCGGAACCCCAATCTGTTCGCCACCCCTTATCTGGCA 462
Db 137 LeuSerGluProPro-----ProTyrProAlaAlaLeuProProPro-----LeuAla 152

QY 463 CAGGCTGCCGGAGGGACCTTTGGCCCTCTCGAGCTCCG-----GCCGTGGAG 513
Db 153 ProProAlaVal----GlyPro---AlaProGlyGlnAlaProAspSerSerAspProGlu 170

QY 514 GGACCTGTGTCAGGAGCTCGGAGCGCGGAGCGGCCACCCCG-----GAGCGGACA 564
Db 171 GlyProAlaAlaGlyThrArgSerArgArgAlaArgSerProAlaAspAspSerGlyPro 190

QY 565 GACGAGATCGGCACATTACCGCTCGGCAGCTACGCGCTCCGCCACCGGGGGCCAAATG 624
Db 191 AspSerThrValIleLeuProLeuArgAlaIleGlyProProAlaGluProAsnGlyLeu 210

QY 625 CAGCCCTCCAGTATTCGCTTCTTCTTCGACATCTCTATATTTGGAACCTAACCAT 684
Db 211 ValProLeuGlnTyrTrpProPheSerSerAlaAspLeuTyrAsnTrpLysSerAsnHis 230

QY 685 CCCCTTTCTCGGAGGATCCCAACGCTCACGGGTTGGTGAGTCCCTTATCTCTCT 744
Db 231 ProSerPheSerGluAsnProAlaGlyLeuThrGlyLeuLeuGluSerLeuMetPheSer 250

QY 745 CACAGCCTACTTGGGATGATTGTCAACAGCTGTGTCAGACACTCTTTCACAACCGAGGAG 804
Db 251 HisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnIleLeuPheThrThrGluGlu 270

QY 805 CGAGAGAGATCTTATTAGAGGCTAGAAAAATGTTCTCGGGCCGACGGCGACCCACG 864
Db 271 ArgGluArgIleLeuLeuGluAlaArgLysAsnValLeuGlyAspAsnGlyAlaProThr 290

QY 865 CGGTTCAAAAATCAGATTGACATGGGATTTCCCTTAACTCGCCCGGTTGGGACTACAAC 924
Db 291 GlnLeuGluAsnLeuIleAsnGluAlaPheProLeuAsnArgProGlnTrpAspTyrAsn 310

QY 925 ACGCTGAAGGTAGGGAGAGCTTGAATAATCTATCGCAGGCTCTGGTGGCGGGTCTCCGG 984
Db 311 ThrAlaAlaGlyArgGluLeuLeuValTyrArgArgThrLeuValAlaGlyLeuLys 330

QY 985 GGGCCTCAGACGGCCCACTAATTTGGCTTAAGTAAAGAGATGATGAGGACCGAAT 1044
Db 331 GlyAlaAlaArgArgProThrAsnLeuAlaLysValArgGluValLeuGlnGlyProAla 350

QY 1045 GAACCCCTCTGTTTTTCTTGAGAGCTCTTGGAGCCCTTCAGGGCGGTACACCCCTTTT 1104
Db 351 GluProProSerValPheLeuGluArgLeuMetGluAlaTyrArgArgTyrThrProPhe 370

QY 1105 GATCCCACCTCAGAGGGCCCAAAAGCCTCAGTGGCTTTTGGCCTTTATAGCACAGTCAGCC 1164

```

```

Db 371 AspProSerGluGluGlyGlnGlnAlaAlaValAlaThrAlaPheThrGlyGlnSerAla 390
QY 1165 TTGATATTAGAAAGAGCTTCAGAGACTGGAAGGTTTACAGGAGGCTGAGTTACGTGAT 1224
Db 391 ProAspIleLysLysLysLeuGlnArgLeuGluGlyLeuGlnAspTyrSerLeuGlnAsp 410
QY 1225 CTAGTGAAGGAGGAGAGAAAGTATATTACAAAAGGGAGACAGAGAAGAAAGGAAACAA 1284
Db 411 LeuValArgGluAlaGluLysValTyrHisLysArgGluThrGluGluLysArgGlnGlu 430
QY 1285 AGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1344
Db 431 ArgGluLysLysGluAlaGluGluArgGluArgArgAspArgArgGlnGluLysAsn 450
QY 1345 TTGACTAAGATCTTGGCTGCGAGTGGTT---GAAGGGAAGAACCAATACGGAAGAGAGA 1401
Db 451 LeuThrArgIleLeuAlaAlaValValSerGly----- 462
QY 1402 GATTTTAGGAAAATTAGGTCTAGGCTCAGGCACTCAGGAGACCTGGGCAATAGGACC--- 1458
Db 463 -----GlySerGlyAspArgArgThrGlyAsnLeuGlyAsnArgAlaArg 477
QY 1459 -----CCACTCGACAAGGACCAATGTGCATATTTGTAAGAA 1494
Db 478 LysThrProArgAspGlyArgProProLeuAspLysAspGlnCysAlaTyrCysLysGlu 497
QY 1495 AGAGGACACTGG 1506
Db 498 LysGlyHisTrp 501

RESULT 3
FOMVM7
gag polyprotein - baboon endogenous virus (strain M7)
N;Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein
C;Species: baboon endogenous virus
A;Note: host Papio sp. (baboon)
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C;Accession: J02060
R;Kato, S.; Matsuo, K.; Nishimura, N.; Takahashi, N.; Takano, T.
Jpn. J. Genet. 62, 127-137, 1987
A;Title: The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric gene
A;Reference number: J02060
A;Accession: J02060
A;Molecule type: DNA
A;Residues: 1-537 <KAT>
A;Cross-references: UNIPROT:P03341; UNIPARC:UPI000012AFEE; GB:D10032; GB:D00088; MID:g2
C;Comment: This protein is synthesized as a gag-pol polyprotein.
C;Genetics:
A;Gene: gag
C;Superfamily: mammalian retrovirus gag polyprotein 1
C;Keywords: core protein; inner coat protein; nucleoprotein; polyprotein
F;1-112/Product: inner coat protein p12 #status predicted <INP>
F;113-226/Product: core protein p15 #status predicted <CPP>
F;227-477/Product: core shell protein p30 #status predicted <CSP>
F;478-537/Product: nucleoprotein p10 #status predicted <NPP>

Alignment Scores:
Pred. No.: 1,49e-93 Length: 537
Score: 1477.50 Matches: 293
Percent Similarity: 66.4% Conservative: 80
Best Local Similarity: 52.1% Mismatches: 126
Query Match: 51.9% Indels: 63
DB: 1 Gaps: 10

US-10-723-552-3_copy_585_2156 (1-1572) x FOMVM7 (1-537)
QY 1 ATCGGACAGAGCGTGACGACCCCTCTTAGTTTTCACCTCTCGACCATTTGGAGCTGAAGTTAAA 60
Db 1 MetGlyGlnThrLeuThrThrProLeuSerLeuThrLeuThrHisPheSerAspValArg 20
QY 61 TCCAGGGCTCATAATTTGTGTCAGTTAAGAGGAGACCTTGGCAGACTTTCGTGTC 120
:::|||||

```

```
Db      21  AlaArgAlaHisAsnLeuSerValGlyValArgLysGlyArgTrpGlnThrPheCysSer  40
Qy      121 TCTGAATGCCCGCATTCGATGTTGGATGGCCATCAGAGGGAGCTTTAAATTCGAGATT 180
Db      41  SerGluTrpProThrLeuHisValGlyTrpProArgAspGlyThrPheAspLeuSerVal  60
Qy      181 ATCTCGGCTGTAAAGCAGATTATTTTTCAGACTCGACCCCGCTCTCATCCCGATCAGGAG 240
Db      61  IleuGlnValLysThrLysValMetAspProGlyProHisGlyHisProAspGlnVal  80
Qy      241 CCTATATCTTACGTGGCAAGATTTCGACAGGATCTCCGCCATGGTGTAAACCATGG 300
Db      81  AlatyrlleileThrTrpGluAspLeuValArgAsnProProProTrpValLysProPhe  100
Qy      301 CTGAATAGCCAGAAGCCAGGTCGCCGAATTCGTGCTCTTCGAGAGAAACAAACAC 360
Db      101 LeuHisThrProSerThrSerLysSerThrLeuLeuAlaLeuGluValProLysAsnArg  120
Qy      361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAG----- 402
Db      121 ThrLeuAspProProLysPro-----ValLeuProAspGluSerGlnGlnAspLeu  137
Qy      403 -----ATTGAGGAGCCACCG-----GTTGGCCGGAAACCCCAA 435
Db      138 LeuPheGlnAspProLeuProHisProProHisAsnProLeuLeuGluProProProTyr  157
Qy      436 TCTGTTCCCCACCCCTTATCTGGCAGAGGTCCGGGAGGGACCTTTGGC----- 489
Db      158 AsnSerProSerProProValLeuSerProValSerProThrThrProSerAlaProThr  177
Qy      490 -----CCTCCTGGAGCTCCGGCG----- 507
Db      178 ProSerSerLeuValSerSerThrProProSerSerProAlaProProGluLeuThr  197
Qy      508 -----GTGGAGGACCTGCTCGAGGACTCGGAGCGCGGAGCGCCACCCCG 555
Db      198 ProArgThrProProGlnThrProArgLeuArgLeuArgAlaGluGlyGlnAspGly  217
Qy      556 GAGCGGACAGACGAGATCGACATTACCGTGGCGCAGTACGGCCCTCCACACCGGG 615
Db      218 ProSerThrTrpGlnSerSerLeuPheProLeuArgThrValAsn----- 232
Qy      616 GCACCAATGCCCTTCTCGAGGATCCCAACCCCTCAGGGGTTGGTGAGTCCCTT 735
Db      232 -----ArgThrIleGlnTrpProPheSerAlaSerAspLeuTyrAsnTrpLys  249
Qy      676 ACTAACCATCCCTTCTCGAGGATCCCAACCCCTCAGGGGTTGGTGAGTCCCTT 735
Db      250 ThrHisAsnProSerPheSerGlnAspProGlnAlaLeuThrSerLeuIleGluSerIle  269
Qy      736 ATGTTCTCTCACCGCTTCTCGAGGATGATTGTCAACAGCTGCTGCACACACTCTTCA 795
Db      270 LeuLeuThrHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnValLeuLeuThr  289
Qy      796 ACCGAGGACGACGAGATTCTATTAGAGGCTAGAAAAATGTTCTCGGGCGCGACGG 855
Db      290 ThrGluGluArgGlnArgLeuLeuLeuGluAlaArgLysAsnValProGlyProGlyGly  309
Qy      856 GCACCCGCGGTTGCAAAATGAGATTGACATGGGATTTCCTTAACTCGCCCCGGTTGG 915
Db      310 LeuProThrGlnLeuProAsnGluIleAspGluGlyPheProLeuThrArgProAspTrp  329
Qy      916 GACTACACACGGCTGAAGGTAGGAGAGCTTGAATAATCTATCGCCAGGCTCTGTGGCG 975
Db      330 AspTyrGluThrAlaProGlyArgGluSerLeuArgIleTyrArgGlnAlaLeuLeuAla  349
Qy      976 GGTCTCCGGGGCGCTCAAGACGGCCCTAATTTGGCTTAAGGTAAAGAGATGATGAG 1035
Db      350 GlyLeuLysGlyValGlyLysArgProThrAsnLeuAlaLysValArgThrIleThrGln  369
Qy      1036 GGACCGAATGAACCCCTCTCTTTTCTTGAGAGGCTCTTGGAAGCCCTTCAGCGCGTAC 1095
Db      370 GlyLysAspGluSerProAlaAlaPheMetGluArgLeuLeuGluGlyPheArgMetTyr  389
```

```
Qy      1096 ACCCTTTTGCATCCACCTCAGAGGCCCAAAAAGCCTCAGTGGCTTTGGCTTTATAGGA 1155
Db      390 ThrProPheAspProGluAlaProGluHisLysAlaThrValAlaMetSerPheIleasp  409
Qy      1156 CAGTACGCTTGGATATTGAAAGAACCTTCAGAGACTGGAAGGTTACAGGAGGCTGAG 1215
Db      410 GlnAlaAlaLeuAspIleLysGlyLysLeuGlnArgLeuAspGlyIleGlnThrHisGly  429
Qy      1216 TTACCTCATCTAGTGAAGGAGGAGAAAGTATATTACAAAAGGAGACAGAGAAGAA 1275
Db      430 LeuGlnGluLeuValArgGluAlaGluLysValTyrAsnLysArgGluThrProGluGlu  449
Qy      1276 AGGGAACAAAGAAAGAGAGAGAAAGAGAGAAAGGAGAGAAACCTTAATAACGGCAA 1335
Db      450 ArgGluAlaArgLeuIleLysGluGlnGluArgGluAspArgAspArgLysArg  469
Qy      1336 GAGAAGAATTGACTAAGATCTTGGCTGCGAGTGTGGAAGGAAAAGCAATACCGAAAGA 1395
Db      470 AspLysHisLeuThrLysIleLeuAlaAlaValVal-----ThrGluLys  484
Qy      1396 GAGAGACATTTAGGAAAATTAGTCCAGGCCCTAGACAGTCCAGGAACTGGGCAATAGG 1455
Db      485 -----ArgAlaGly-----LysSerGlyGluThrArgArgArg  495
Qy      1456 ACCCCACTCGACAGGACCAATGTGCATATTGTAAAGAAAGAGGACACTGGCGCAAGAAC 1515
Db      496 ProLysValAspLysAspGlnCysAlaTyrCysLysGluArgGlyHisTrpIleLysasp  515
Qy      1516 TGCCCCCAAGAAG-----GGAAACAAAGGACCAAGGATCTTAGCTCTAGAGAAGAT 1566
Db      516 CysProLysArgProArgAspGlnLysLysProAlaProValLeuThrLeuGlyGluAsp  535
Qy      1567 AAAGAT 1572
Db      536 SerGlu 537

RESULT 4
R35474
gag polyprotein - Friend murine leukemia virus (strain PVC-211)
N;Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein p12; Friend murine leukemia virus
C;Species: Friend murine leukemia virus
A;Variety: strain PVC-211
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C;Accession: S35474
R;Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.; Masuda, M.
Nucleic Acids Res. 20, 3249, 1992
A;Title: Complete nucleotide sequence of a neuropathogenic variant of Friend murine leukemia virus
A;Reference number: S35474; MUID:92319660; PMID:1620621
A;Accession: S35474
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-538 <REM>
A;Cross-references: UNIPROT:P26805; UNIPARC:UPI000012B02C; EMBL:M93134; NID:9331898; PID:9331898
A;Experimental source: strain PVC-211
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992
C;Genetics:
A;Gene: gag
C;Superfamily: mammalian retrovirus gag polyprotein I
C;Keywords: core protein; inner coat protein; nucleoprotein; polyprotein

Alignment Scores:
Pred. No.: 4.14e-87 Length: 538
Score: 1383.50 Matches: 285
Percent Similarity: 62.3% Conservative: 69
Best Local Similarity: 50.2% Mismatches: 137
Query Match: 48.6% Indels: 77
DB: 2 Gaps: 14

US-10-723-552-3_COPY_585_2156 (1-1572) x S35474 (1-538)

Qy      1  ATGGACGACGCGTGACGACCCCTCTTAGTTGACTCTCGACCATTCGACCTGAAGTTAA 60
Db      1  ATGGACGACGCGTGACGACCCCTCTTAGTTGACTCTCGACCATTCGACCTGAAGTTAA 60
```

Db 1 MetGlyGlnThrAlaThrThrProLeuSerLeuThrLeuAspHisTrpLysAspValGlu 20  
Qy 61 TCCAGGGCTCAATAATTTGTCAAGTTAAGAGGACCTTGGCAGACTTCTGTGTC 120  
Db 21 ArgThrAlaHisAsnGlnSerValGluValArgLysArgArgTrpValThrPheCysSer 40  
Qy 121 TCTGAATGGCCGACATTCGATGTTGGATGGCCATCAGAGGGACCTTTAATCTGAGATT 180  
Db 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnProAspIle 60  
Qy 181 ATCTCGGCTGTAAAGCAGTTATTTTTCAGACTGACCCGCTCTCATCCGATCAGGAG 240  
Db 61 IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80  
Qy 241 CCTATATCCTACGTGGCAAGATTGGCAGAGGATCCTCCGCCATGGTTAAACCATGG 300  
Db 81 ProTyrlleValThrTrpGluAlaLeuAlaValAspProProTrpValLysProPhe 100  
Qy 301 CTGAATTAAGCCAAAGACCGAGTCCCGGAATTCCTGGCTCTTGGAGAGAAAAAACAAC 360  
Db 101 ValHis-----ProLysProProLeuLeu-----ProPro 111  
Qy 361 TCGGCTGAAAAGTCAAGCCCTCTCTCT-----CATATCTAC 396  
Db 112 SerAlaProSerLeuProGluProProLeuSerThrProProGlnSerSerLeuTyr 131  
Qy 397 CCCGAGATTGAGGACCA---CCGGCTTGGCCGGAACCCCAATCTGTTCCTC----- 444  
Db 132 ProAlaLeuThrSerProLeuAsnThrLysProArgProGlnValLeuProAspSerGly 151  
Qy 445 -----CCACCCCTTATCTGGCAGAGGT----- 468  
Db 152 GlyProLeuIleAspLeuLeuThrGluAspProProProTyrArgAspProGlyProPro 171  
Qy 469 -----GCCGGAGGGACCTTTCCTCT-----CCTGAGCTCCGGCGGTG 510  
Db 172 SerProAspGlyLysGlyAspSerSerGlyGluValAlaProThrGluGlyAlaProAspSer 191  
Qy 511 GAGGGACCTGCTGAGGAGTCCGAGCGCCGAGGGCGCCACCCCGAGCGACAGACGAG 570  
Db 192 SerProMetValSerArgLeuArgGlyArgGluProProValAlaAspSerThrThr 211  
Qy 571 ATCGGACATTCAGCTGCGCACGTACGGCCCTCCACACCGGGGGGCCAAATTCAGCCCC 630  
Db 212 SerGlnAlaPheProLeuArgLeuGlyGly-----AsnGlyGln----- 224  
Qy 631 CTCAGTATTGGCCCTTTCTTCGACAGATCTATAATTGGAAAACTAACCATCCCTT 690  
Db 225 PheGlnTyrTrpProPheSerSerSerAspLeuTyrAsnTrpLysAsnAsnAsnProSer 244  
Qy 691 TTCTCGGAGGATCCCAACGCTCACCGGGTGTGGAGTCCCTTATGTTCTCTCACCAG 750  
Db 245 PheSerGluAspProGlyLysLeuThrAlaLeuIleGluSerValLeuLeuThrHisGln 264  
Qy 751 CCTACTTGGGATGATTCTCAACAGCTCTGCAGACACTCTTCAACCCGAGGAGCGAGAG 810  
Db 265 ProThrTrpAspAspCysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGluGluLysGln 284  
Qy 811 AGAATTCATTAGAGGTAGAAAAATGTTCTTGGGCCGACCGGGCCACCCACCGCGTGT 870  
Db 285 ArgValLeuLeuGluAlaArgLysAlaValArgGlyGluAspGlyArgProThrGlnLeu 304  
Qy 871 CAATAATCAGATTGACATGGATTTCCTTAACTCGCCCGCGGTGGGACTACACACGGCT 930  
Db 305 ProAsnAspIleAsnAspAlaPheProLeuGluArgProAspTrpAspTyrAsnThrGln 324  
Qy 931 GAAGGTAGGAGAGCTGAAAATCTATCGCAGGCTCTGGTGGCGGTCTCCGGGGCGCC 990  
Db 325 ArgGlyArgAsnHisLeuValHisTyrArgGlnLeuLeuLeuAlaGlyLeuGlnAsnAla 344  
Qy 991 TCAAGACGGCCCACTAATTTGGCTTAAGTAAGAGATGATGACGAGCCGAATGAACCC 1050  
Db 345 GlyArgSerProThrAsnLeuAlaLysValLysGlyIleThrGlnGlyProAsnGluSer 364

Qy 1051 CCCTCTGTTTTTCTTGAGAGCTCTTGGAAAGCCTTCAGGGGTACACCCCTTTTGATCCC 1110  
Db 365 ProSerAlaPheLeuLeuGluArgLeuLysGluAlaTyrArgArgTyrThrProTyrAspPro 384  
Qy 1111 ACCTCAGAGCCCAAAAAGCCTCAGTGGCTTGGCCCTTTATAGGACAGTACGCTTGAT 1170  
Db 385 GluAspProGlyGlnGluThrAsnValSerMetSerPheIleTrpGlnSerAlaProAsp 404  
Qy 1171 ATTAGAAGAGCTTCAGAGACTCGAAGGGTTACAGGAGGCTGAGTTACGTGATCTAGTG 1230  
Db 405 IleGlyArgLysLeuGluArgLeuGluAspLysAsnLysThrLeuGlyAspLeuVal 424  
Qy 1231 AAGGAGCGCAGAAAGTATATTACAAAAGGGAGAGACAGAAGAAAGGAAACAAAGAAA 1290  
Db 425 ArgGluAlaGluLysIlePheAsnLysArgGluThrProGluGluArgGluGluArgVal 444  
Qy 1291 GAGAGAAAGAGAGAAAGGAGGAAAGA----- 1320  
Db 445 ArgArgGluThrGluGluLysGluGluArgArgAlaGluAspGluArgGluLys 464  
Qy 1321 ---CGTAATAAAGCGCAAGAGAGAAATTTGACTTAAGATCTTGGCTGCAGTGGTTGAAGG 1377  
Db 465 GluArgAspArgArgArgHisArgGluMetSerLysLeuLeuAlaThrValValSerGly 484  
Qy 1378 AAAAGCAATACGGAAAGAGAGAGAGATTTTAGGAAATTAGGTCCAGCCCTAGACAGTCA 1437  
Db 485 Gln-----ArgGlnAspArgGlnGlyGlyGluArg--- 494  
Qy 1438 GGGAACTGGGCAATAGACCCCATCGACAAGCAATGTGCATATTGTAAGAAAGA 1497  
Db 495 -----ArgArgProGlnLeuAspHisAspGlnCysAlaTyrCysLysGluLys 510  
Qy 1498 GGCACCTGGCAAGGAACTGCCCCAGAG-----GGAACAAGGACCAAGG----- 1545  
Db 511 GlyHisTrpAlaArgAspCysProLysLysProArgGlyProArgGlyProArgProGln 530  
Qy 1546 -----ATCCTAGCTCTAGAAGAA 1563  
Db 531 AlaSerLeuLeuThrLeuAspAsp 538

## RESULT 5

S70394

gag polyprotein - Friend murine leukemia virus (strain FB29)

N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleopro

C:Species: Friend murine leukemia virus

A:Variety: strain FB29

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004

C:Accession: S70394

R;Perryman, S.; Nishio, J.; Chesebro, B.

Nucleic Acids Res. 19, 6950, 1991

A;Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.

A;Reference number: S70393; MUID:92107687; PMID:1762923

A;Accession: S70394

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: genomic RNA

A;Residues: 1-538 &lt;PER&gt;

A;Cross-references: UNIPROT:P26806; UNIPARC:UP1000012802B; EMBL:Z11128; NID:G61547; PID

A;Experimental source: strain FB29

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C;Genetics:

A;Gene: gag

C;Superfamily: mammalian retrovirus gag polyprotein I

C;Keywords: core protein; inner coat protein; nucleoprotein; polyprotein

## Alignment Scores:

Pred. No.:	5,67e-87	Length:	538
Score:	1381.50	Matches:	285
Percent Similarity:	62.9%	Conservative:	72
Best Local Similarity:	50.2%	Mismatches:	134
Query Match:	48.5%	Indels:	77
DB:	2	Gaps:	15

US-10-723-552-3 COPY 585 2156 (1-1572) X S70394 (1-538)

Qy	1	ATGGGACAGCGGTGACGACCCCTCTTAGTTTGACTCTCGACATTGGACTGAAGTAAA	60
Db	1	MetGlyGlnAlaValThrProLeuSerLeuThrLeuAspHisTrpLysAspValGlu	20
Qy	61	TCCAGGGCTCATAAATTTGTCAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTCTGTGTGC	120
Db	21	ArgThrAlaHisAsnLeuSerValGluValArgLysArgArgTrpValThrPheCysSer	40
Qy	121	TCTGAATGGCGGACAAATTCGATTTGGATGGCCATCAGAGGGGACCTTTTAATTTCTGAGATT	180
Db	41	AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnProAspIle	60
Qy	181	ATCTCGGTGGTGTAAAGCAGTATTATTTTCAGACTGGACCGCGTCTCATCCCATCAGGAG	240
Db	61	IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal	80
Qy	241	CCCTATATCTTACGTGCAAGATTGGCAGAGGATCTCCGCCATGGGTTAAACCATGG	300
Db	81	ProTyrIleValThrTyrGluAlaIleAlaValAspProProProTrpValArgProPhe	100
Qy	301	CTGAATAAGCCAGAAAGCCAGGTCCCCGAATTTGGCTCTTGGAGAGAAAAACAACAC	360
Db	101	ValHis---ProLysProPro-----LeuSerLeu-----ProPro	111
Qy	361	TCGGCTGAAAAAGTCACAGCCCTCTCCT-----CATATCTAC	396
Db	112	SerAlaProSerLeuProProGluProProLeuSerThrProGlnSerSerLeuTyr	131
Qy	397	CCCGAGATTGAGGAGCCA---CCGGCTTGGCGGAACCCCAATCTGTTCC	444
Db	132	ProAlaLeuThrSerProLeuAsnThrLysProArgProGlnValLeuProAspSerGly	151
Qy	445	-----CCACCCCTTATCTGGCACAGGT-----	468
Db	152	GlyProLeuIleAspLeuLeuThrGluAspProProProTyrArgAspProGlyProPro	171
Qy	469	-----GCCGCGAGGGGACCTTTGGCCCT---CCTGGAGCTCCGCGCGTG	510
Db	172	SerProAspGlyAsnGlyAspSerGlyGluValAlaProThrGluGlyAlaProAspPro	191
Qy	511	GAGGACCTGTCGACGAGGACTCGGACCGGAGGGGGCCACCCGGAGCGGACAGACGAG	570
Db	192	SerProMetValSerArgLeuArgLysGlyLysGluProProValAlaAspSerThrThr	211
Qy	571	ATCGGCATTAACCGCTGCGCAGCTACGGCCCTCCACACCGGGGGCCAAATTGCAGCCC	630
Db	212	SerGlnAlaPheProLeuArgLeuGlyGly-----AsnGlyGln-----	224
Qy	631	CTCCAGTATGGCCCTTTTCTCTGAGATCTCTATAATTGGAATACTTAACCATCCCCCT	690
Db	225	TyrGlnTyrTrpPheSerSerSerAspLeuTyrAsnTrpLysAsnAsnAsnProSer	244
Qy	691	TTCTCGGAGGATCCCAACGCTCAGCGGTGGTGGAGTCCCTTATGTTCTCTCACACG	750
Db	245	PheSerGluAspProAlaLysLeuThrAlaLeuIleGluSerValLeuLeuThrHisGln	264
Qy	751	CCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTCAACACCCGAGGACGAGAG	810
Db	265	ProThrTrpAspAspCysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGluGluLysGln	284
Qy	811	AGAATTCATTAGAGGCTAGAAAAAATGTTCTTGGGGCCGACGGCGACCCACGCGGTG	870
Db	285	ArgValLeuLeuGluAlaArgLysAlaValArgGlyGluAspGlyArgProThrGlnLeu	304
Qy	871	CAAAATGAGATTGACATGGGATTTCCCTTAACCTCGCCCGGTTGGGACTTACAACACGGCT	930
Db	305	ProAsnAspIleAsnAspAlaPheProLeuGluArgProAspTrpAspIlyrAsnThrGln	324
Qy	931	GAAGGTAGGGAGAGCTTGAAATCTATCCCAAGGCTCTGTCGCGGCTCTCCGGCGCC	990
Db	325	ArgGlyArgAsnHisLeuValHisTyrArgGlnLeuLeuLeuLeuAlaGlyLeuGlnAsnAla	344

991	QY	TC	AAGACGGCCCACTAATTTGGCTTAAGTAAAGAGAGTGTATGACGAGCGACCGAATGAACCC	1050
345	Db	Gly	ArgSerProThrAsnLeuAlaLysValLysGlyIleThrGlnGlyProAsnGluSer	364
1051	QY	CC	CTCTGTGTTTTCTTCGAGAGCCTCTTGGAAAGCCTTCAGGCGGTACACCCCTTTGTATGCC	1110
365	Db	Pro	SerAlaPheLeuGluArgLeuLysGluAlaIyrArgArgTyrThrProTyrAspPro	384
1111	QY	AC	CTCAGAGCCCAAAAGCCTCAGTGGCTTTGGCCCTTTATAGACACAGTCAGCCTTGGAT	1170
385	Db	Glu	AspProGlnGlnThrAsnValAlaMetSerPheIleTrpGlnSerAlaProAsp	404
1171	QY	AT	TAGAAGAAGCTTCAGACAGCTGGAGGGTTACAGGAGCGCTGAGTTACGTGATCTAGTG	1230
405	Db	Ile	GlyArgLysLeuGluArgLeuGluAspLysSerLysThrLeuGlyAspLeuVal	424
1231	QY	AAG	GACGACAGAAAAGTATATTACAAAGGGAGACAGAGAAGAAAGGGAACAAAGAAA	1290
425	Db	Arg	GluAlaGluLysIlePheAsnLysArgGluThrProGluGluArgGluGluArgIle	444
1291	QY	GAG	AGAGAAGACAGGAAGGGAGGAAGA-----	1320
445	Db	Arg	ArgGluThrGluGluLysGluGluArgArgAlaGluAspValGlnArgGluLys	464
1321	QY	---	CGTAATAACGGCAAGAGAAATTTTCACATAAGATCTTGGCTGCAGTGGTTGAAGGG	1377
465	Db	Glu	ArgAspArgArgGlnHisArgGluMetSerLysLeuLeuAlaThrValSerGly	484
1378	QY	AA	AGCAATACGGAAAGAGAGAGAGATTTTAGGAAATATTAGTCAGGCCCTAGACAGTCA	1437
485	Db	Gln	-----ArgGlnAspArgGlnGlyGlyGluArg---	494
1438	QY	GG	GAACCTGGCAATAGGACCCCACTCGACAGGACCAATGTGCATATTGTAAAGAAGA	1497
495	Db	---	ArgArgProGlnLeuAspHisAspGlnCysAlaTyrCysLysGluLys	510
1498	QY	GG	CACCTGGCAGGAACCTCCCAAGAAG-----GGAAACAAAGGACCAAGG-----	1545
511	Db	Gly	HisTrpAlaArgAspCysProLysLysProArgGlyProArgGlyProArgProGln	530
1546	QY	----	ATCCTAGCTCTAGAAGAA	1563
531	Db	Ala	SerLeuLeuThrLeuAspAsp	538

## RESULT 6

**FROM**

gag polyprotein - Moloney murine sarcoma virus  
N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein p12  
C:Species: Moloney murine sarcoma virus  
C:Date: 18-Dec-1981 #sequence revision 18-Dec-1981 #text\_change 09-Jul-2004  
C:Accession: A94261; B00645; A03932  
R:Reddy, E.P.; Smith, M.J.; Aaronson, S.A.  
Science 214, 445-450, 1981  
A:Title: Complete nucleotide sequence and organization of the Moloney murine sarcoma virus  
A:Reference number: A94261; MUID:82039559; PMID:6170110  
A:Accession: A94261  
A:Molecule type: genomic RNA  
A:Residues: 1-538 <RED>  
A:Cross-references: UNIPROT:P03334; UNIPARC:UPI000012B032; GB:J02266; NID:g331973; PIDN:J02266  
A:Experimental source: provirus  
R:Van Beveren, C.; van Straaten, F.; Galleshaw, J.A.; Verma, I.M.  
Cell 27, 97-108, 1981  
A:Title: Nucleotide sequence of the genome of a murine sarcoma virus.  
A:Reference number: A00645; MUID:82115347; PMID:6173134  
A:Accession: B00645  
A:Molecule type: DNA  
A:Residues: 1-518, 'K', 520-538 <VAN>  
A:Cross-references: UNIPARC:UPI0000100AB5; GB:V01185; GB:J02263; NID:g61647; PIDN:CAA245454  
A:Experimental source: clone 124, circular  
C:Genetics:  
A:Gene: gag  
C:Superfamily: mammalian retrovirus gag polyprotein I

C;Keywords: polypeptide  
F;2-131/Product: core protein p15 #status predicted <p15>  
F;132-215/Product: inner coat protein p12 #status predicted <p12>  
F;216-478/Product: core shell protein p30 #status predicted <p30>  
F;479-534/Product: nucleoprotein p10 #status predicted <p10>

Alignment Scores:  
Pred. NO.: 7,78e-87 Length: 538  
Score: 1379.50 Matches: 280  
Percent Similarity: 61.8% Conservative: 77  
Best Local Similarity: 48.4% Mismatches: 124  
Query Match: 48.4% Indels: 97  
DB: 1 Gaps: 13

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x FOMVM (1-538)

QY 1 ATGGGACAGCGGTGACGACCCCTCTTAGTTTGGACTCTCGACCATTTGGACTGAAGTAAA 60  
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpLysAspValGlu 20  
QY 61 TCCAGGGCTCATAATTTGTCAGTTTTCAGGTTAAGAGGACCTTGGCAGACTTTCTGTGTC 120  
Db 21 ArgLeuAlaHisAsnGlnSerValAspValLysLysArgArgTrpValThrPheCysSer 40  
QY 121 TCTGAATGGCCGACATTCGATGTTGGATGGCCATCAGAGGGACCTTTAATCTGAGATT 180  
Db 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60  
QY 181 ATCTGGCTGTAAAGCAGTTATTATTTTCAGACTGACCGGCTCTCATCCGATCAGAG 240  
Db 61 IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80  
QY 241 CCCTATATCTTACGTGGCAAGATTGGCAGAGGATCTCCGCCATCGGTTAAACCATGG 300  
Db 81 ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProProTrpValLysProPhe 100  
QY 301 CTGAATGAAGCAAGAACCGAGTCCCGCAATTCCTGGCTCTTGGAGAGAGAAAAACAAC 360  
Db 101 ValHis---ProLysProProProProLeuLeuProSer----- 112  
QY 361 TCGGCTGAAAAAGTCAAGCCCTCTCATATCTACCCGAGATTGAGAGCCACCG--- 417  
Db 113 -----AlaProSerLeuProLeuGluProProLeuSerThrProProGln 127  
QY 418 -----GCTTGGCCGGCAACCCCAA----- 435  
Db 128 SerSerLeuTyrProAlaLeuThrProSerLeuGlyAlaLysProLysProGlnValLeu 147  
QY 436 -----TCTGTTCCCCCAACCCCTTATCTGGCA 462  
Db 148 SerAspSerGlyProLeuIleAspLeuLeuThrGluAspProProProTyr----- 165  
QY 463 CAGGGTCCCGAGGGGACCTTTGGCCCTCCT----- 495  
Db 166 -----ArgAspProArgProProProSerAspArgAspGlyAspSerGlyGlu 181  
QY 496 GGAGCTCCGGCGGTGGAGGACCT-----GCTGAGGGACTCTCGAGCCGG 540  
Db 182 AlaThrProAlaGlyGluAlaProAspProSerProMetAlaSerArgLeuArgGlyArg 201  
QY 541 AGGGGCCACCCCGGAGCGGACAGACAGATCCGCACTTACCGTGGCAGCATCGGC 600  
Db 202 ArgGluProProValAlaAspSerThrThrSerGlnAlaPheProLeuArgThrGlyGly 221  
QY 601 CCTCCCACCGGGGGGCAATTCGACCCCTCCAGTATGGCCCTTTCTTCTGACAGAT 660  
Db 222 -----AenGlyGln-----LeuGlnTyrTrpProPheSerSerAsp 234  
QY 661 CTCTATATTTGGAATACTAACCATCCCTTTCTCGAGGAGTCCCAACGCCCTCAGCGGG 720  
Db 235 LeuTyrAsnTrpLysAsnAsnAspProSerPheSerGluAspProGlyLysLeuThrAla 254  
QY 721 TTGGTGGAGTCCCTTATGTTCTCTCACCAGCCTACTTGGGATGATTGTCAACAGCTGCTG 780

Db 255 LeuIleGluSerValLeuIleThrHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 274  
QY 781 CAGACACTTCTTCAACCGAGGACGAGAGAAATTCATTAGAGGCTAGAAAAAATCTT 840  
Db 275 GlyThrLeuLeuThrGlyGluGluLysGlnArgValLeuLeuGluAlaArgLysAlaVal 294  
QY 841 CTGGGGCCGACGGGCGACCCACCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA 900  
Db 295 ArgGlyAspAspGlyArgProThrGlnLeuProAsnGluValAspAlaPheProLeu 314  
QY 901 ACTGCCCCGGTTGGGACTCAACACCGCTCAAGGTAGGAGAGCTTGAATAATCTATCGC 960  
Db 315 GluArgProAspTrpGluTyrThrGlnAlaGlyArgAsnHisLeuValHisTyrArg 334  
QY 961 CAGGCTCTGGTGGGGGCTCTCCGGGGCGCTCAAGACGCGCCACTTAATTTGCGTAAGTA 1020  
Db 335 GlnLeuLeuIleAlaGlyLeuGlnAsnAlaGlyArgSerProThrAsnLeuAlaLysVal 354  
QY 1021 AGAAGTGTATGACGAGGACCGAATGAACCCCTCTGTTTTTCTTGAGAGGCTCTTGAA 1080  
Db 355 LysGlyIleThrGlnGlyProAsnGluSerProSerAlaPheLeuGluArgLeuLysGlu 374  
QY 1081 GCCTTACGGGCTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAAGCCCTCAGTGC 1140  
Db 375 AlaTyrArgArgTyrThrProTyrAspProGluAspProGlyGlnGluThrAsnValSer 394  
QY 1141 TTGSCCTTTATAGACAGTCAAGCTTGCATATTAGAAAGAGCTTCAGAGACTCGAAGG 1200  
Db 395 MetSerPheIleTrpGlnSerAlaProAspIleGlyArgLysLeuGluArgLeuGluAsp 414  
QY 1201 TTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAGGACAGAGAAGTATATTACAAAG 1260  
Db 415 LeuArgAsnLysThrLeuGlyAspLeuValArgGluAlaGluArgIlePheAsnLysArg 434  
QY 1261 GAGACAGAAGAAGAGGGAACAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db 435 GluThrProGluArgGluArgGluArgIleArgArgGluArgGluArgLysGluArg 454  
QY 1321 -----CGTAATAACCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1347  
Db 455 ArgArgThrGluAspGluGlnLysGluLysGluArgAspArgArgHisArgGluMet 474  
QY 1348 ACTAAGATCTTGGCTGAGTGGTTGAAGGAGAAAGCAATACGGAAGAGAGAGAGAGATTT 1407  
Db 475 SerArgLeuLeuAlaThrValValSerGlyGlnArgGlnAspArgGlnGlyGluArg 494  
QY 1408 AGGAAATATTAGTCAAGCCCTCAGACATCAGGAGAACCTGGGCAATAGAGACCCCACTG 1467  
Db 495 Arg-----ArgSerGlnLeuAsp 500  
QY 1468 AAGGACCAATGTGCATATTCTAAAGAGAGGACACTGGGCAAGGAGTCCCAAG--- 1524  
Db 501 CysAspGlnCysThrTyrCysGluGluGlnGlyHisTrpAlaLysAspCysProArg 520  
QY 1525 ---AAGGAAACAAAGGACCAAGG-----ATCCTAGCTCTAGAGAA 1563  
Db 521 ProArgGlyProArgGlyProArgProGlnThrSerLeuLeuThrLeuAspAsp 538

RESULT 7

T10532

gag-pol polyprotein precursor - feline leukemia virus (strain FelV-FAIDS)

N;Contains: core protein p10; core protein p12; core protein p15; core protein p27; en

C;Species: feline leukemia virus

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004

C;Accession: T10532

R;Donahue, P.R.; Hoover, E.A.; Beltz, G.A.; Riedel, N.; Hirsch, V.M.; Overbaugh, J.; Mu

J. Virol. 62, 722-731, 1988

A;Title: Strong sequence conservation among horizontally transmissible, minimally patho

A;Reference number: Z17078; MUID:88119207; PMID:2828667

A;Accession: T10532

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: mRNA





A;Title: Nucleotide sequence of Moloney murine leukaemia virus.

A;Reference number: A93265; MUID:82035843; PMID:6169994

A;Accession: A03930

A;Molecule type: Genomic RNA

A;Residues: 1-538 <SH1>

A;Cross-references: UNIPROT:P03332; UNIPARC:UPI000017424

A;Experimental source: clone pMLV-1

C;Comment: This protein is synthesized as a gag-pol polyprotein.

C;Genetics:

A;Gene: gag

C;Superfamily: mammalian retrovirus gag polyprotein 1

C;Keywords: core protein, inner coat protein; nucleoprotein; polyprotein

F;2-131/Product: core protein p15 #status predicted <C15>

F;132-215/Product: inner coat protein p12 #status predicted <C12>

F;216-478/Product: core shell protein p30 #status predicted <C30>

F;479-534/Product: nucleoprotein p10 #status predicted <C10>

Alignment Scores:

Pred. No.:	1-25e-86	Length:	538
Score:	1376.50	Matches:	283
Percent Similarity:	62.0%	Conservative:	68
Best Local Similarity:	50.0%	Mismatches:	142
Query Match:	48.3%	Indels:	73
DB:	1	Gaps:	13

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x FOMVIM (1-538)

```

QY 1 ATGGGACAGAGCGGTGACGACCCCTCTTAGTTTGAAGTTCGACCATTCGACTGAAGTTAAA 60
DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20

QY 61 TCCAGGCTCATATTTGTCTAGTTTCAAGTAAAGAGGACCTTGGCAGACTTTCTGTGTC 120
DB 21 ArgileAlaHisAsnGlnSerValAspValLysArgTrpValThrPheCysSer 40

QY 121 TCTGAATGGCGGACATTCGATGTGGATGGCATCAGAGGACCTTTAATTCGAGATT 180
DB 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60

QY 181 ATCTGGCTGTAAAGCAGTATTTTTCAGACTGACGACCGGCTCTCATCCGATCAGGAG 240
DB 61 IleThrGlnValLysIleValPheSerProGlyHisGlyHisProAspGlnVal 80

QY 241 CCTTATATCTTACGTGGCAAGATTTGGCAGAGGATCTCCGCGCATCGGTAAACCATGG 300
DB 81 ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProTrpValLysProPhe 100

QY 301 CTGAATAGCAAGAACCCAGGTCCCGAATTCCTGGCTCTTGAGAGAGAAAAACAACAC 360
DB 101 ValHis---ProLysProProProProLeuProProSerAlaProSerLeuProLeuGlu 119

QY 361 TCGGCTGAAAAGTCAAGCCCTCTCTATATCTACCCGAGATTGAGGAGCCA---CCG 417
DB 120 ProProArgSerThrProProArgSerSerLeuTyrProAlaLeuThrProSerLeuGly 139

QY 418 GCTTGGCCGGAACCCCAA-----GGAGCTCCGCGGTGAGGACCT-----TCT 438
DB 140 AlaLysProLysProGlnValLeuSerAspSerGlyGlyProLeuIleAspLeuLeuThr 159

QY 439 GTTCCCCCAACCCCTTATCTGGCACAGGTGCGCGAGGGAGACCTTTGCCCCCTCCT--- 495
DB 160 GluAspProProTyr-----ArgAspProArgProProProSer 173

QY 496 -----GGAGCTCCGCGGTGAGGACCT----- 519
DB 174 AspArgAspGlyAsnGlyGlyGluAlaThrProAlaGlyGluAlaProAspProSerPro 193

QY 520 ---GCTGCAGGAGCTCGAGCGGAGGGCGGCCACCCCGGAGCGGACAGACGAGATCGCG 576
DB 194 MetAlaSerArgLeuArgGlyArgGluProProValAlaAspSerThrThrSerGln 213

QY 577 ACATTACCGTCGCGACGTAAGCCCTTCCACACCCGCGGGGCGCAATTCGAGCCCTCCAG 636
DB 577 MetAlaSerArgLeuArgGlyArgGluProProValAlaAspSerThrThrSerGln 213

```

RESULT 9

FOMVGV

gag polyprotein - AKV murine leukemia virus

```

Db 214 AlaPheProLeuArgAlaGlyGly-----AsnGlyGln-----LeuGln 226
QY 637 TATTGGCCCTTTCTTCTCGAGACTCTATAATTGGAAAACCTAACCATCCCCCTTCTCG 696
Db 227 TyrTrpProPheSerSerAspLeuTrpAsnTrpLysAsnAsnProSerPheSer 246
QY 697 GAGGATCCCAACGCCCTCACGGGTTGGTGGAGTCCCTTATGTTCTCTCACGAGCTACT 756
Db 247 GluAspProGlyLysLeuThrAlaLeuIleGluSerValLeuIleThrHisGlnProThr 266
QY 757 TGGGATGATTCGACAGCTGCTCCAGACTCTTCAACCCGAGGAGCGAGAGAAATT 816
Db 267 TrpAspAspCysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGluGluLysGlnArgVal 286
QY 817 CTATTAGAGCTAGAAAAATGTTCTCGGCGCGACCGGCGGACCCAGCGGTTCCCAAT 876
Db 287 LeuLeuGluAlaArgLysAlaValArgGlyAspAspGlyArgProThrGlnLeuProAsn 306
QY 877 GAGATTGACATGGGATTTCCTTAACCTCGCCCGGTTGGGACTACAAACCGCTGAAGGT 936
Db 307 GluValAspAlaAlaPheProLeuGluArgProAspTrpPheTyrThrThrGlnAlaGly 326
QY 937 AGGAGAGCTTGAATAATCTATCGCCAGGCTCTGTGGCGGCTCTCCGGGCGGCTCAAGA 996
Db 327 ArgAsnHisLeuValHisTyrArgGlnLeuLeuAlaGlyLeuGlnAsnAlaGlyArg 346
QY 997 CGGCCCACTAATTTGGCTAAGTAAGTAAGTATGATCGAGGACCGAATGAACCCCTCT 1056
Db 347 SerProThrAsnLeuAlaLysValLysGlyIleThrGlnGlyProAsnGluSerProSer 366
QY 1057 GTTTTCTTTCAGAGGCTCTTGGAAAGCCTTCAGGCGGTACACCCCTTTTGTATCCACCTCA 1116
Db 367 AlaPheLeuGluArgLeuLysGluAlaTyrArgArgTyrThrProTyrAspProGluAsp 386
QY 1117 GAGCCCAAAAGCCTCAGTGGCTTGGCTTATAGGACAGTCAGCTTGGATATTAGA 1176
Db 387 ProGlyGlnGluThrAsnValSerMetSerPheIleTrpGlnSerAlaProAspIleGly 406
QY 1177 AAGAAGCTTCAGAGACTGGAAGGTTACAGGAGGCTGAGTACGTGATCTAGTGAAGGAG 1236
Db 407 ArgLysLeuGlyArgLeuGluAspLeuLysAsnLysThrLeuGlyAspLeuValArgGlu 426
QY 1237 GCAGAGAAATGATTATCAAAAGGAGACAGAGAAGAAAGGAAACAAAGAGAGAGA 1296
Db 427 AlaGluLysIlePheAsnLysArgGluThrProGluGluArgGluGluArgIleArgArg 446
QY 1297 GAAAGAGAGAAAGGAGGAGAAAGA-----CGT 1323
Db 447 GluThrGluGluLysGluGluArgArgThrGluAspGluGlnLysGluLysGluArg 466
QY 1324 AATAAGCGCAAGAGAAATTTGACTAGATCTTGGCTCGAGTGGTTGAAGGAGAAAGC 1383
Db 467 AspArgArgHisArgGluMetSerLysLeuLeuAlaThrValValSerGlyGlnLys 486
QY 1384 AATACGAAAGAGAGAGAGATTTTAGGAAAATTAGGTACGCGCTCAGAGTACAGGAAC 1443
Db 487 Gln-----AspArgGlnGlyGlyGlu 493
QY 1444 CTGGGCAATAGGACCCCACTCGAAGAGCAATGTGCATATTGTAAAGAAAGAGGACAC 1503
Db 494 ---ArgArgArgSerGlnLeuAspArgAspGlnCysAlaTyrCysLysGluLysGlyHis 512
QY 1504 TGGGCAAGGACTGCCCCCAGAG-----GGAAACAAAGGACCAAG----- 1545
Db 513 TrpAlaLysAspCysProLysLysProArgGlyProArgGlyProArgProGlnThrSer 532
QY 1546 ATCTAGCTCTAGAGAA 1563
Db 533 LeuLeuThrLeuAspAsp 538

```

N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein  
 C:Species: AKV murine leukemia virus  
 A:Note: host Mus sp. (mouse)  
 C:Date: 03-Aug-1984 #sequence\_revision 27-Nov-1985 #text\_change 09-Jul-2004  
 C:Accession: A03933  
 R:Herr, W.

J. Virol. 49, 471-478, 1984  
 A:Title: Nucleotide sequence of AKV murine leukemia virus.  
 A:Reference number: A92995; MUID:84115072; PMID:6319746  
 A:Accession: A03933  
 A:Molecule type: genomic RNA  
 A:Residues: 1-537 <HER>  
 A:Cross-references: UNIPROT:P03336; UNIPARC:UPI000012B025; GB:J01998; GB:J01999; GB:K000  
 C:Comment: This protein is synthesized as a gag-pol polyprotein.  
 C:Genetics:

A:Gene: gag  
 C:Superfamily: mammalian retrovirus gag polyprotein I  
 C:Keywords: core protein; inner coat protein; nucleoprotein; polyprotein  
 F:1-129/Product: core protein p15 #status predicted <C15>  
 F:130-214/Product: inner coat protein p12 #status predicted <C12>  
 F:215-477/Product: core shell protein p30 #status predicted <C30>  
 F:478-537/Product: nucleoprotein p10 #status predicted <C10>

Alignment Scores:  
 Pred. No.: 2,98e-86 Length: 537  
 Score: 1371.00 Matches: 282  
 Percent Similarity: 61.6% Conservative: 72  
 Best Local Similarity: 49.0% Mismatches: 129  
 Query Match: 48.1% Indels: 92  
 DB: 1 Gaps: 14

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x FOMGV (1-537)

Qy 1 ATGGGACAGCGGTGACGACCCCTCTTAGTTGACTCTCGACCAATTGGACTGAAGTTAAA 60  
 Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGluHisTrpGluAspValGln 20  
 Qy 61 TCCAGGGCTCATATTTCTCAGTTTTCAGTTTGAAGAGGACCTTGGCAGACTTTCTGTGTC 120  
 Db 21 ArgileAlaSerAsnGlnSerValAspValIlylsYsArgArgTrpValThrPheCysSer 40  
 Qy 121 TCTGAATGCCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAAATTCGAGATT 180  
 Db 41 AlaGluTrpProThrPheGlyValGlyTrpProGlnAspGlyThrPheAsnLeuAspIle 60  
 Qy 181 ATCTCGCTGTTAAAGCAGTTATTTTTCAGACTGAGCCCGGCTCTCATCCCGATCAGAG 240  
 Db 61 IleLeuGlnValYsSerIysValPheSerProGlyProHisGlyHisProAspGlnVal 80  
 Qy 241 CCCTATATCCTTACGTGGCAAGATTTCGAGAGGATCTCCGCCATGGTGGTTAAACCATGG 300  
 Db 81 ProTyrIleValThrTrpGluAlaIleAlaTyrGluProProProTrpValYsProPhe 100  
 Qy 301 CTGAATAGCCAAAGAACCCAGGTCCCGCAATTCCTGGCTCTTGGAGAGAAAAACAAACAC 360  
 Db 101 ValSerProIysLeuSerProSerProThrAlaProIleLeu-----Pro 115  
 Qy 361 TCGGCTGAAAAAGTCAAGCCCTCTCTCAT-----ATCTACCCCGAGATTGAGAGGCA 414  
 Db 116 SerGlyProSerThrGlnProProProGlnProGlnProGlnProGlnProGlnProGln 134  
 Qy 415 CCGGCTTGGCCGGAACCCCA-----435  
 Db 135 SerIleIysProArgProSerIysProGlnValLeuSerAspAsnGlyProLeuIle 154  
 Qy 436 -----TCTGTTCCCGCCACCCCTTATCTGGCAGCAGGT-----468  
 Db 155 AspLeuLeuSerGluAspProProProProTyrGlyGlyGlnGlyLeuSerSerAspGly 174  
 Qy 469 -----GCCGCGAGGGAGCCCTTTTGGCCCTCTCTGGAGCTCCGCGGTG 510  
 Db 175 AspGlyAspArgGluAlaThrSerThrSerGluIleProAlaProSerProIleVal 194

Qy 511 -----GAGGGACCTGCTCAGGAGACTCGGAGCCGAGGCGGCC 549  
 Db 195 SerArgLeuArgGlyLysArgAspProProAlaAlaAspSerThrThrSerArg-----212  
 Qy 550 ACCCGGAGCGGACAGACGAGATCGACATTACCGCTCGGACAGTACGGCCCTCCACACA 609  
 Db 213 -----AlaPheProLeuArgLeuGlyGly-----220  
 Qy 610 CCGGGGGCCCAATTGTCAGCCCCCTCCAGTATTGGCCCTTTCTCTCGAGATCTCTATAAT 669  
 Db 221 ---AsnGlyGln-----LeuGlnTyrTrpProPheSerSerAspLeuTyrAsn 236  
 Qy 670 TGGAAAACTAACCATCCCTTTCTCGAGGATCCCAACGCCCTCAGCGGGTTCGTGGAG 729  
 Db 237 TrpIysAsnAsnAsnProSerPheSerGluAspProGlyLysLeuThrAlaLeuIleGlu 256  
 Qy 730 TCCCTTATGTTCTCTCCAGCCCTTACCTGGGATGATTGTCAACAGCTGCTCGACACTC 789  
 Db 257 SerValLeuThrThrHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlyThrLeu 276  
 Qy 790 TTCAACACCGGAGCGAGAGAGATTTCTATTAGAGCTAGAAAAATGTTCTCTGGGCCC 849  
 Db 277 LeuThrGlyGluLysGlnArgValLeuLeuGluAlaArgLysAlaValArgGlyAsn 296  
 Qy 850 GACGGGCGACCCAGCGGTTGCAAAATCAGATTGACATGGGATTTCCCTTAACTCGCCCC 909  
 Db 297 AspGlyArgProThrGlnLeuProAsnGluValAlaAspAlaPheProLeuGluArgPro 316  
 Qy 910 GGTGGGACTACAACCGGTGAAGGTAGGAGAGCTTGAATAATCTATCGCAGGCTCTG 969  
 Db 317 AspTrpAspTyrThrThrGlnArgGlyArgAsnHisLeuValLeuTyrArgGlnLeuLeu 336  
 Qy 970 GTGGCGGCTCCCGGCGCTCAAGACGCCCACTAATTTGGCTTAAGGTAAGAGAGTG 1029  
 Db 337 LeuAlaGlyLeuGlnAsnAlaGlyArgSerProThrAsnLeuAlaLysValIleGlyIle 356  
 Qy 1030 ATGACGAGCGACCGAATGAACCCCTCTGTTTCTTCAGAGGCTCTTGGAAAGCCCTCAGG 1089  
 Db 357 ThrGlnGlyProAsnGluSerProSerAlaPheLeuGluAlaGluLysGluAlaTyrArg 376  
 Qy 1090 CGGTACACCCCTTTGATCCACCTCAGAGGCCCAAAAGCCTCAGTGGCTTTGGCCCTT 1149  
 Db 377 ArgTyrThrProTyrAspProGluAspProGlyGlnGluThrAsnValSerMetSerPhe 396  
 Qy 1150 ATAGACAGCTCAGCTTGGATATTAGAAGAGCTTCAGAGCTGAGAGGCTTACAGGAG 1209  
 Db 397 IleTrpGlnSerAlaProAspIleGlyArgLysLeuGluArgLeuGluAspLeuYsSer 416  
 Qy 1210 GCTGAGTTACGTGATCTAGTGAAGGAGCAGAGAAAGTATATTACAAAGGGAGACAGAA 1269  
 Db 417 LysThrLeuGlyAspLeuValArgGluAlaGluArgIlePheAsnLysArgGluThrPro 436  
 Qy 1270 GAAGAAAGGGAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 Db 437 GluGluArgGluGluArgValArgArgGluThrGluGluLysGluArgArgArgAla 456  
 Qy 1321 -----CGTAATAACCGGCAAGAGAGAAATTCGACTAAGATC 1356  
 Db 457 GluGluGlnLysGluLysGluArgAspArgArgHisArgGluMetSerIysLeu 476  
 Qy 1357 TTGGCTGCGAGTGGTTGAAGGGAAAAAGCAATACGGAAAGAGAGAGAGATTTTAGGAAAATT 1416  
 Db 477 LeuAlaThrValValSerGlyGln-----ArgArgProGlnLeuAspIysAspGln 487  
 Qy 1417 AGGTACGCCCTAGACAGCTCAGGGAACCTGGGCAATAGGACCCCACTCGACAGAGCAAA 1476  
 Db 488 ArgGlnGlyGlyGluArg-----ArgArgProGlnLeuAspIysAspGln 502  
 Qy 1477 TGTCTATTCTTAAGAAAGAGGACACTGGCAAGGAACCTCCCAAGAG-----GGA 1530  
 Db 503 CysAlaTyrCysLysGluLysGlyHisTrpAlaLysAspCysProLysLysProArgGly 522  
 Qy 1531 AACAAAGACCAAGG-----ATCCTAGCTCTAGAAGAA 1563

Db 523 ProArgGlyProArgProGlnThrSerLeuLeuThrLeuAspAsp 537  
RESULT 10  
FOMVRV  
gag polyprotein - radiation murine leukemia virus  
N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein  
C:Species: radiation murine leukemia virus  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 09-Jul-2004  
C:Accession: A26183  
R:Merregaert, J.; Janowski, M.; Reddy, E.P.  
Virolgy 158, 88-102, 1987  
A:Title: Nucleotide sequence of a radiation leukemia virus genome.  
A:Reference number: A94362; MUID:87207680; PMID:3033897  
A:Accession: A26183  
A:Molecule type: DNA  
A:Residues: 1-537 <MER>  
A:Cross-references: UNIPROT:P11269; UNIPARC:UPI000012B02E; GB:K03363; GB:M18449; NID:933  
C:Genetics:  
A:Gene: gag  
C:Superfamily: mammalian retrovirus gag polyprotein I  
C:Keywords: core protein; inner coat protein; nucleoprotein; polyprotein  
F:1-129/Product: core protein p15 #status predicted <p15>  
F:130-214/Product: inner coat protein p12 #status predicted <p12>  
F:215-477/Product: core shell protein p30 #status predicted <p30>  
F:478-537/Product: nucleoprotein p10 #status predicted <p10>  
Alignment Scores:  
Pred. No.: 8.98e-86 Length: 537  
Score: 1364.00 Matches: 280  
Percent Similarity: 62.6% Conservative: 75  
Best Local Similarity: 49.4% Mismatches: 136  
Query Match: 47.9% Indels: 76  
DB: 1 Gaps: 12  
US-10-723-552-3\_COPY\_585\_2156 (1-1572) x FOMVRV (1-537)  
QY 1 ATGGGACAGCGGTGACGACCCCTCTTAGTTGACTCTCGACCAATGGAGCTGAAGTAAA 60  
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGluHisTrpGlyAspValGln 20  
QY 61 TCAGGGCTCATATTTGTCAGTTTCAGTTAAGAGGACCTTGGCAGACTTCTGTGTC 120  
Db 21 ArgIleAlaSerAsnGlnSerValGluValLysLysArgArgValThrPheCysPro 40  
QY 121 TCTGAATGGCGGACATTCGATGTTGGATGGCCATCAGAGGGAGCTTTAATCTGAGATT 180  
Db 41 AlaGluTrpProThrPheAspValGlyTrpProGlnAspGlyThrPheAsnLeuAspIle 60  
QY 181 ATCTGGCTGTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 240  
Db 61 IleLeuGlnValLysSerLysValPheSerProGlyProHisGlyHisProAspGlnVal 80  
QY 241 CCCTATATCTAGTCGCAAGATTGGCAGAGATCTCCGCCATCGGTTAAACCATGG 300  
Db 81 ProTyrIleValThrTrpGluAlaIleAlaTyrGluProSerTrpValLysProPhe 100  
QY 301 CTGAATAAGCAAGAAAGCCAGGTCCCGAATTCGCTCTGGAGAGAAAAACAACAC 360  
Db 101 ValSerProLysLeuSerLeuSerProThrAlaProIleLeu-----Pro 115  
QY 361 TCGGCTGAAAAGTCAAGCCCTCTCTCAT-----ATCTACCCCGAG 402  
Db 116 SerGlyProSerThrGlnProProArgSerAlaLeuTyrProAlaLeuThrProSer 135  
QY 403 ATTGAGAGCACCGGCTTGGCGGAACCCCAATCT-----438  
Db 136 IleLysProArgProSerLysProGlnValLeuSerAsnGlyGlyProLeuLeuAsp 155  
QY 439 -----GTTCCCAACCCCTTATCTGGCAGCAGGTCGCGGAGGAGACCTTGGC 489  
Db 156 LeuLeuThrGluAspProProProTyr-----GlyGluGlnGlyProSerSer 171

QY 490 CCTCTCGA-----GCTCCGCGGTGGAG 513  
Db 172 ProAspGlyAspGlyAspArgGluGluAlaThrTyrThrSerGluIleProAlaProSer 191  
QY 514 GGACTCTGCGAGGACTCGAGCGGAGGGGCCACCCCGAGCGGACGAGCAGATC 573  
Db 192 ProMetValSerArgLeuArgGlyLysArgAspProProAlaAlaAspSerThrThrSer 211  
QY 574 GCGACATTACCGTCCGACGTCAGCGCCCTCCACACCGGGGGGCCAATTGCGAGCCCTC 633  
Db 212 ArgAlaPheProLeuArgLeuGlyGly-----AsnGlyGln-----Leu 224  
QY 634 CAGTATTGGCCCTTTCTTCGACAGATCTTATAATTGGAATAACCATCCCTTTC 693  
Db 225 GlnTyrTrpProPheSerSerSerAspLeuTyrAsnTrpLysAsnAsnProSerPhe 244  
QY 694 TCGAGGATCCCAACCCCTCAGGGGTTGGTGGAGTCCCTTATGTTCTCTCAGCAGCT 753  
Db 245 SerGluAspProGlyLysLeuThrAlaLeuIleGluSerValLeuThrThrHisGlnPro 264  
QY 754 ACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTCAACCGAGGAGCAGAGAGA 813  
Db 265 ThrTrpAspAspCysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGluGlyGlnArg 284  
QY 814 ATTCTATTAGAGCTAGAAAAATGTTCTCGGGCCGAGCGGCGACCCAGCGGTTCAG 873  
Db 285 ValLeuLeuGluAlaArgLysAlaValArgGlyAsnAspGlyArgProThrGlnLeuPro 304  
QY 874 AATGAGATTGACATGGGATTTCCCTTAACCTCGCCCGGTTGGGACTCAACACCGCTGAA 933  
Db 305 AsnGluValAsnSerAlaPheProLeuGluArgProAspTrpAspTyrThrThrProGlu 324  
QY 934 GGTAGGAGAGCTTGAATACTATCGCCAGCTCTGTGGCGGTCTCCGGGGCGCTCA 993  
Db 325 GlyArgAsnHisLeuValLeuTyrArgGlnLeuLeuAlaGlyLeuGlnAsnAlaGly 344  
QY 994 AGACGGCCCACTAATTTGGCTAAGAGAAATCATGACGAGGACCGAATGAACCCCCC 1053  
Db 345 ArgSerProThrAsnLeuAlaLysGlyLysGlyIleThrGlnGlyProAsnGluSerPro 364  
QY 1054 TCTGTTTCTTTCAGAGGCTCTTGGAGGCTTTCAGGGGTACACCCCTTTTGTATCCAC 1113  
Db 365 SerAlaPheLeuGluArgLeuLysGluAlaTyrArgArgTyrThrProTyrAspProGlu 384  
QY 1114 TCAGAGCCCAAGAGCTCAGTGGCTTTCGCTTTATAGACAGTCTCAGCTTGGATATT 1173  
Db 385 AspHisGlyGlnGluThrSerValSerMetSerPheIleTrpGlnSerAlaProAspIle 404  
QY 1174 AGAAGAGCTTCAGAGACTGGAAGGTTTACAGAGGCTCAGTTACGTGATCTAGTGAAG 1233  
Db 405 GlyArgLysLeuGluArgLeuGluAspLeuLysSerLysThrLeuArgAspLeuValArg 424  
QY 1234 GAGCGAGAGAAATATATTACAAAGGAGACAGAGAGAAAGGGAACAAAGAAAGAG 1293  
Db 425 GluAlaGluLysIlePheAsnLysArgGluThrProGluGluArgGluArgPheArg 444  
QY 1294 AGAAGAGAGAGAAAGGAGGAGAAAGA-----1320  
Db 445 ArgGluThrGluGluAsnGluGluArgArgAlaGluAspGluGlnArgGluLysGlu 464  
QY 1321 CGTAATAAAGCGCAAGAGAAATTTGACTTAAGATCTTGGCTCGAGTGGTTGAAGGAAA 1380  
Db 465 ArgAspArgArgGlnArgGluMetSerLysLeuLeuAlaThrValValThrGlyGln 484  
QY 1381 AGCAATACGGAAGAGAGAGATTTTAGAAATTTAGGTACGCGCTAGACAGTCAAGG 1440  
Db 485 -----ArgGlnAspArgGlnGlyGlyGluArg-----493  
QY 1441 AACCTGGCAATAGGACCCCACTCGACAGGACCAATGTCATATTGTAAGAAAGAGGA 1500  
Db 494 -----LysArgProGlnLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 510  
QY 1501 CACTGGGCAAGAACTGCCCCAAGAAG-----GGAAACAAGGACCAAGG-----1545



```

116 SerGlyProSerThrGlnProProArgSerAlaLeuTyrProAlaphethrProSer 135
      |||:::|||||||
403 ATTGAGGAGCCACGGGTTCGCCGAACCCCCAATCT----- 438
136 IleLysProArgProSerLysProGlnValLeuSerAspGlycylProLeuilleAsp 155
      :::::|||||||
439 -----GTTCCCCACCCTTATCTGGCACAGGTCGCCGAGGAGCCTTTGCC 489
      |||||
156 LeuLeuThrGluaspProProPofry-----GlyGluGlnGlyProSerSer 171
      |||||
490 CCTCCTCGA-----GCTCCGCCTGGAG 513
      |||||
```

## RESULT 12

[illegible]

245 SerGluaspProGlyLysLeuThrAlaLeuIleGluSerValLeuThrThrHisGlnPro 264

754 ACTTGGGATGATTGTCAACAGCTGCTGCAGACACTCTTCAACAACCGAGGAGCGAGAGAGA 813  
265 ThrTriaAspCysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGluGluGlySerGlnArg 284  
814 ATTCTATTAGAGGCTAGAAAAAATGTTCTCTGGGGCCGACGGGCGACCCACCGCGGTTCGAA 873  
::: ::::

Pred. No.:	2,71e-85	Length:	537
Score:	1357.00	Matches:	279
Percent Similarity:	62.6%	Conservative:	76
Best Local Similarity:	49.2%	Mismatches:	136
Query Match:	47.6%	Indels:	76
DB:	1	Gaps:	13

03-10-723-332-3 \_COFI\_365\_2136 (1-13/2) X FOMVMB (1-33/)

[illegible]

```
QY 1381 AGCAATACGGNAAGAGAGAGAGATTTTAGGAAATTTAGTTCAGGCCCTTAGACAGTCAGGG 1440
Db 485 -----ArgGlnAspArgGlnGlyGlyGluArg----- 493
QY 1441 AACCTGGGCAATAGGACCCCACTGCACAGGACCAATGTGCATATTTGTAAGAAAGAGGA 1500
Db 494 -----ArgArgProGlnLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 510
QY 1501 CACTGGGCAAGAACTGCCCAAGAAG-----CGAAACAAGACCAAG----- 1545
Db 511 HisTrpAlaLysAspCysProLysLysProArgGlyProArgGlyProArgProGlnThr 530
QY 1546 ---ATCTAGCTCTAGAGAA 1563
Db 531 SerLeuLeuThrLeuAspAsp 537

RESULT 13
FOMVME
gag polyprotein - murine leukemia virus (strain DEF27)
N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein
C:Species: murine leukemia virus
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C:Accession: B40416
R:Chattopadhyay, S.K.; Sengupta, D.N.; Fredrickson, T.N.; Morse III, H.C.; Hartley, J.W.
J. Virol. 65, 4232-4241, 1991
A:Title: Characteristics and contributions of defective, ecotropic, and mink cell focus-
A:Reference number: A40416; MUID:91303677; PMID:1649328
A:Accession: B40416
A:Molecule type: DNA
A:Residues: 1-536 <CHA>
A:Cross-references: UNIPROT:P29168; UNIPARC:UPI000012B028; GB:M64096
C:Genetics:
A:Gene: gag
A:Superfamily: mammalian retrovirus gag polyprotein I
C:Keywords: core protein; inner coat protein; nucleoprotein; polyprotein
F:1-129/Product: core protein p15 #status predicted <CP>
F:130-213/Product: inner coat protein p12 #status predicted <LCP>
F:214-476/Product: core shell protein p30 #status predicted <CSP>
F:477-536/Product: nucleoprotein p10 #status predicted <NP1>

Alignment Scores:
Pred. No.: 1,216-84 Length: 536
Score: 1347.50 Matches: 282
Percent Similarity: 62.3% Conservative: 69
Best Local Similarity: 50.1% Mismatches: 127
Query Match: 47.3% Indels: 85
DB: 1 Gaps: 15

US-10-723-552-3_COPY_585_2156 (1-1572) x FOMVME (1-536)
QY 1 ATGGACAGAGCGGTGACGACCCCTCTAGTTTGACTCTCGACCAATTGACCTGAAGTTAA 60
Db 1 MetGlyGlnThrIleThrProLeuSerLeuThrLeuGluHisTrpArgAspValGln 20
QY 61 TCCAGGGCTCATAAATTTGTCAGTTTGAAGGGACCTTGGCAGACTTTCGTGTC 120
Db 21 CysIleAlaSerAsnGlnSerValAspValLysArgArgTrpValThrPheCysSer 40
QY 121 TCTGAATGCCGACATTTCGATGCTGGATGGCCATCAGAGGGGACCTTTTAATTTCTGAGATT 180
Db 41 ValGluTrpProSerPheAspValGlyTrpProLeuAspGlyThrPheAsnLeuAspIle 60
QY 181 ATCTCTGGCTGTTAAAGCAGTATATTTTTCAGACTGGACCGGGCTCTCATCCCGATCAGGAG 240
Db 61 IleLeuGlnValLysSerLysValPheCysProGlyProHisGlyHisProAspGlnVal 80
QY 241 CCTATATCTTACGTGGCAAGATTGGCAGAGGATCTCCGCCATGGTGGTTAAACATGG 300
Db 81 ProTyrIleValThrTrpGluAlaLeuAlaTyrHisProProTrpValLysProPhe 100
QY 301 CTGAAT---AGCCA-----AGAAAGCCAGGTCCC----- 327
```

```
Db 101 ValSerProLysProPheProLeuSerThrLeuProPheSerProProGlyProSerAla 120
QY 328 -----CGAATTTCTGGCTCTTGGAGAGAAAAACAACAC 360
Db 121 HisProProSerArgSerAspLeuTyrThrAlaLeuIleProSerIleLysThrLysPro 140
QY 361 TCGGCTGAAAAAGTCAAGCCCTCT-----CCTCATATCTACCCCGAGATTGAGGAG 411
Db 141 ProLysSerArgValLeuProThrAsnGlyGlyProLeuIleAspLeuLeuThrGluAsn 160
QY 412 CCA-----CCGCTTGGCCGCGAACCACCAATCTGTT 441
Db 161 ProProAsnLeuGlyGluGlnGlyProProLeuProLysGlyProValLysLysArgArg 180
QY 442 CCCCACCCCTTATCTGGCACAGGTGCGCGAGGGGACCTTTCGCCCTCTCTGGAGCT 501
Db 181 ProProProPro-----ArgTyrSerProProGly--- 190
QY 502 CCGCGGTGGAGGACCTGCTGCAGGACTCGAGGCGGAGGGCGCCACCCCGAGCGG 561
Db 191 ProMetVal-----SerArgLeuArgGlyAsnArgAspProProAla 204
QY 562 ACAGACGAGATCGGACA-----TTACCGCTCGGCACGTACGGCCCTCCACACCGGG 615
Db 205 AlaAspSerThrThrSerArgAlaPheProLeuArgLeuGlyGly-----Asn 220
QY 616 GGCCAAATGAGCCCTCCAGTATTCGCCCTTCTTCAGATCTCTATAATTGAAA 675
Db 221 GlyGln-----LeuGlnTyrTrpProPheSerSerSerAspLeuTyrAsnTrpLys 237
QY 676 ACTAACCATCCCCCTTTCGAGGATCCCAACCGCTCACGGGGTGTGGAGTCCCT 735
Db 238 AsnAsnAsnProSerPheSerGluAspProGlyLysLeuThrAlaLeuIleGluSerVal 257
QY 736 ATGTTCTCTCACGACCTACTTGGGATGTGTCAACAGCTGCTGCAGACACTCTTCA 795
Db 258 LeuThrThrHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlyThrLeuLeuThr 277
QY 796 ACCGAGGAGCGAGAGAAATCTATTAGAGCTAGAAAAATGTTCTCGGGCGCGAGGG 855
Db 278 GlyGluGluLysGlnArgValLeuLeuGluAlaArgLysAlaValArgGlyAsnAspGly 297
QY 856 CGACCCACGCGGTGTCAAATGAGATTGACATGGGATTTCCCTTAACTCGCCCGGTGG 915
Db 298 ArgProThrGlnLeuProAsnGluValAsnSerAlaPheProLeuGluArgProAspTrp 317
QY 916 GACTACAACACGGCTGAAGGTGAGGAGCTTGAATAATCTATCCCGAGGCTCTCGTGGCG 975
Db 318 AspTyrThrThrProGluGlyArgAsnHisLeuValLeuTyrArgGlnLeuLeuAla 337
QY 976 GGTCTCCGGGCGCTCAAGACGGCCCACTAATTTGGCTAAGTAAGAGAGTGTATGCAG 1035
Db 338 GlyLeuGlnAsnAlaGlyArgSerProThrAsnLeuAlaLysValLysGlyIleThrGln 357
QY 1036 GGACCGAATGAACCCCTCTGTTTCTTGAGAGGCTCTTGGAGGCTCTCAGCGCGGTAC 1095
Db 358 GlyProSerGluSerProSerAlaPheLeuGluArgLeuLysGluAlaTyrArgArgTyr 377
QY 1096 ACCCTTTTGTATCCCACTCAGAGCGCCCAAAACGCTCAGTGGCTTGGCCCTTATAGGA 1155
Db 378 ThrProTyrAspProGluAspProGlyGlnGluThrAsnValSerMetSerPheIleTrp 397
QY 1156 CAGTCAGCTTGGATATTAGAAAGACTTCAGAGACTGGAGGTTTACAGAGGCTGAG 1215
Db 398 GlnSerAlaProAspIleGlyArgLysLeuGluArgLeuGluAspLeuLysSerLysThr 417
QY 1216 TTACTGTATCTAGTGAAGGCGCAGAGAAATATATTACAAAAGGGGACAGAGAGAA 1275
Db 418 LeuGlyAspLeuValArgGluAlaGluLysIlePheAsnLysArgGluThrProGluGlu 437
QY 1276 AGGGAACAAGAAAGAGAGAGAAAGAGAGAAAGAGAGAAAGAGAGAGAGAGAGAG 1320
Db 438 ArgGluGluAlaGlyIleArgArgGluThrGluGluLysGluGluArgArgAlaGluAsp 457
```





```
QY 1321 -----CCTAATAACGCGAAGAGAAATTTGACTAAGATCTTGGCTGCAGTGGTT 1371
Db 454 GluLysGluArgAspArgArgHisArgGluMetSerLysPheLeuAlaThrValVal 473
QY 1372 GAAGGGAAGCAATACCGAAAGAGAGAGAGATTTTAGGAAATTTAGGTCCAGGCCCTAGA 1431
Db 474 ThrGlyGln-----ArgGlnAspArgGlnGlyGlyGlu 484
QY 1432 CAGTCAGGGAACCTGGGCAATAGGACCCACTCGACAAGGACCAATGTGCATATTTGTA 1491
Db 485 Arg-----ArgArgProGlnLeuAspGluAspGlnCysAlaTyrCysLys 499
QY 1492 GAAGAGAGACTGGGCAAGNACTGCCCCCAAGAG-----GGNAACAAGGACCAAGG 1545
Db 500 GluLysGlyHisTrpAlaLysAspCysProLysLysProArgGlyProArgGlyProArg 519

RESULT 15
FOMVVB
gag polyprotein - baboon endogenous virus
N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprotein
C:Species: baboon endogenous virus strain M7
A:Note: host Papio sp. (baboon)
C:Date: 19-Feb-1984 #sequence_revision 27-Nov-1985 #text_change 09-Jul-2004
C:Accession: A03939
R:Tamura, T.
J. Virol. 47, 137-145, 1983
A:Title: Provirus of M7 baboon endogenous virus: nucleotide sequence of the gag-pol region
A:Reference number: A03939; MUID:83241915; PMID:6408267
A:Accession: A03939
A:Molecule type: DNA
A:Residues: 1-532 <TM>
A:CROSS-references: UNIPROT:P03341; UNIPARC:UPI0000174A29
C:Comment: This protein is synthesized as a gag-pol polyprotein.
C:Genetics:
A:Gene: gag
C:Superfamily: mammalian retrovirus gag polyprotein 1
C:Keywords: core protein; inner coat protein; nucleoprotein; polyprotein
F:1-111/Product: inner coat protein p12 #status predicted <C12>
F:112-218/Product: core protein p15 #status predicted <C15>
F:219-472/Product: core shell protein p30 #status predicted <C30>
F:473-532/Product: nucleoprotein p10 #status predicted <C10>

Alignment Scores:
Pred. No.: 3,29e-82 Length: 532
Score: 1312.00 Matches: 279
Percent Similarity: 64.0% Conservative: 78
Best Local Similarity: 50.0% Mismatches: 141
Query Match: 46.1% Indels: 60
DB: 1 Gaps: 15

US-10-723-552-3_COPY_585_2156 (1-1572) x FOMVVB (1-532)
QY 1 ATGGACAGAGCGTGACGACGCCCTCTTAGTTTGACTCTCGACCATTTGGACTGAAGTTAA 60
Db 1 MetGlyGlnThrLeuThrProLeuSerLeuThrLeuThrHisPheSerAspValArg 20
QY 61 TCCAGGGCTCAATATTCAGTTTCAGTTTAAGAGGGACCTTGGCAGACTTTCTGTGTC 120
Db 21 AlaArgAlaHisAsnLeuSerValGlyValArgLysGlyArgTrpGlnThrPheCysSer 40
QY 121 TCTGAA-----TGCCCGACATTCGATGTTGGATGGCCCATCAGAGGGACCTTTAATTCGAG 177
Db 41 SerGluValHisProSerCysArgValAla-----ArgAspGlyThrPheAspLeuSer 58
QY 178 ATTATCTCGGCTGTTAAAGCAGTATTTTTCAGACTGACCGCGGCTCTCATCCCGATCAG 237
Db 59 ValIleLeuGlnValLysThrLysAspMetAspProGlyProHisGlyHisProValPro 78
QY 238 GAGCCCTATATCTTACGTGGCAAGATTTGGCAGAGGATCTCCGCCATGGGTTAAACCA 297
Db 79 SerGlyTyrIleIleThrTrpValAspLeuAspGlyAsnProProProTrpGlyLysPro 98
```

```
QY 298 TGGCTGAATAAGCAAGCAAGCCAGGTCCTCCCGAATTTCTGGCTCTTGGAGAGAAAAACAA 357
Db 99 PheLeuHisThrProSerThrSerLysSerThrLeuLeuAlaLeuGluValProLysAsn 118
QY 358 CACTCGGCTGAAGAAAGTCAAGCCCTCTCTCATATCTACCCCGAG----- 402
Db 119 ArgThrLeuAspProProLysPro-----ValLeuProAspGluSerGlnGlnAsp 135
QY 403 -----ATTGAGGAGCA-----CCGCTTGGCGGAAACCCCAATCTGTTCCTCCACCCCT 453
Db 136 LeuLeuPheGlnAspProLeuProHisProHisAsnProLeuLeuGluProProPro 155
QY 454 TATCTGCGACAGGTCGCGGAGGGGACCTTTCCTCTCTGAGCTCCGGCGGTG--- 510
Db 156 TyrAsnSerProSerProProThrProSerAlaProThrProSerSerLeuValSer 175
QY 511 -----GAGGACCTGCTCGAGGAGCTCGGACCGGAGGGCGCCACCCCG 555
Db 176 SerSerThrProProSerSerProAlaProProGluLeuThrProArg-----ThrPro 193
QY 556 GAGCGGACAGCAGATC----- 573
Db 194 ProGlnThrProArgLeuArgLeuArgAlaGluGlyGlnAspGlyProPheHisLeu 213
QY 574 -----GCGACATTACCGCTCGCACAGCTACGGCCCTCCACACCGGGGGCCAAATTG 627
Db 214 GlnSerSerLeuPheProLeuArgThrValAsnArgThrValLys-----Arg 229
QY 628 CCCCTCCAGTATTCGCTCTTCTCTGAGACTCTATATAATCGAAACCTAACCATCCC 687
Db 230 ThrIleGlnTyrTrpProPheSerAlaSerAspLeuTyrAsnTrpLysThrHisAsnPro 249
QY 688 CCTTCTCGGAGGATCCCAACCGCTCACGGGGTGTGGAGTCCCTATATTTCTCTCAC 747
Db 250 SerPheSerGlnAspProGlnAlaLeuThrSerLeuIleGluSerIleLeuLeuThrHis 269
QY 748 CAGCTACTCTGGGATGTTGTCAACAGCTGTCGAGACACTCTTTCACAACCGAGAGCGA 807
Db 270 GlnProThrTrpAspAspCysGlnGlnLeuLeuGlnValLeuLeuThrThrGluGluArg 289
QY 808 GAGAGAATTCATTAGAGGCTAGAAAAAATTTCTCTGGGCGCGAGCGGCGACCCACGCGG 867
Db 290 GlnArgValLeuLeuLeuGluAlaArgLysAsnValProGlyProGlyGlyLeuProThrGln 309
QY 868 TTGCAAAATGAGATTGACATGGGATTTCCCTTAACCTCGCCCGGTGGAGCTACACACG 927
Db 310 LeuProAsnGluIleAspGluGlyPheProLeuThrArgProAspTrpAspTyrGluThr 329
QY 928 GCTGAAGGTAGGAGAGCTTGAATAATCTATCCGACAGGCTCTGGTGGCGGGTCTCCGGGCG 987
Db 330 AlaProGlyArgGluSerLeuArgIleTyrArgGlnAlaLeuLeuAlaGlyLeuLysGly 349
QY 988 GCCTCAAGACGGCCCACTAATTTGGCTAAGTAAGTAAGTATGATCGAGGACCGAATGAA 1047
Db 350 AlaGlyLysArgProThrAsnLeuAlaLysValArgThrIleThrGlnGlyLysAspGlu 369
QY 1048 CCCCTCTGTTTCTTTCAGAGGCTCTTGAAGCTTCAGCGGCTTACACCCCTTTTCAT 1107
Db 370 SerProAlaAlaPheMetGluArgLeuLeuGluGlyPheArgMetTyrThrProPheAsp 389
QY 1108 CCCACCTCAGAGGCCCAAAAAGCCTCAGTGGCTTTGGCTTTATAGGACAGTCAGCCCTG 1167
Db 390 ProGluAlaProGluHisLysAlaThrValAlaMetSerPheIleAspGlnAlaLeu 409
QY 1168 GATATTAGAAAGAGCTTTCAGAGACTGGAAGGTTTACAGGAGGCTGAGTTACGTGATCTA 1227
Db 410 AspIleLysGlyLysLeuGlnArgLeuAspGlyIleGlnThrHisGlyLeuGlnGluLeu 429
QY 1228 GTGAAGGCGCAGAGAAAGTATATTACAAAAGGAGACAGAGAGAAAGGAGCAACAGA 1287
Db 430 ValArgGluAlaGluLysValTyrAsnLysArgGluThrProGluGluArgGluAlaArg 449
QY 1288 AAGAGAGAGAAAGAGAGAGAAAGGAGAAAGAGCTATATAACGGAAGAGAGAAATTTG 1347
```



This Page Blank (uspto)

GenCore version 5.1.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 12:52:14 ; Search time 65.4051 Seconds  
(without alignments)  
3391.447 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_585\_2156

Perfect score: 2848

Sequence: 1 ATGGCAGACGGTGACGAC.....CTCTAGAGAGATAAAGAT 1572

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ARSSWEB.spool/US10723552/runat 14022006 125141 12876/app query.fasta\_1  
-DB-Uniprot -QEMT-fastan -SUFFIX=rup -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US10723552 @CGN 1 1 808 @runat 14022006 125141 12876 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -OSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2817	98.9	524	2	O62706_PIG
2	2813	98.8	524	2	O62704_PIG
3	2812	98.7	524	2	O5QG09_9GAMR
4	2774	97.4	524	2	O4VF23_9GAMR
5	2757	96.8	524	2	O6UJ23_9GAMR
6	2747	96.5	524	2	O6T3V4_9GAMR
7	2747	96.5	524	2	O8UWA0_9GAMR
8	2745	96.4	524	2	O90RM0_9GAMR
9	2725	95.7	524	2	O6UJ25_9GAMR
10	2724	95.6	524	2	O6UJ22_9GAMR
11	2705	95.0	664	2	O9N177_PIG
12	2705	95.0	1718	2	O8J4V8_9GAMR
13	2705	95.0	1720	2	O8J4V6_9GAMR
14	2705	95.0	2378	2	O9Q1X4_9GAMR
15	2702	94.9	524	2	O62708_PIG
16	2701	94.8	525	2	O6T3V5_9GAMR

17	2699	94.8	524	2	Q8UMP6_9GAMR	Q8ump6 porcine end
18	2699	94.8	524	2	Q90RL7_9GAMR	Q90rl7 porcine end
19	2699	94.8	2376	2	Q9Q1X3_9GAMR	Q9qlx3 porcine end
20	2698	94.7	524	2	Q8Q6U8_9GAMR	Q8q6u8 porcine end
21	2695	94.6	2376	2	Q9Q1X5_9GAMR	Q9qlx5 porcine end
22	2693	94.6	524	2	Q7ZBT7_9GAMR	Q7zbt7 porcine end
23	2686	94.3	524	2	Q8UM97_9GAMR	Q8um97 porcine end
24	2676	94.0	524	2	Q73504_9GAMR	Q73504 porcine end
25	2419.5	85.0	525	2	Q8Q6U5_9GAMR	Q8q6u5 porcine end
26	1733	60.8	522	2	Q9WLL1_9GAMR	Q9wll1 mus dunni e
27	1733	60.8	622	2	Q89814_9GAMR	Q89814 mus dunni e
28	1718.5	60.3	521	1	Q9TTC2_PHACI	Q9ttc2 phascolat
29	1672.5	58.7	519	1	GAG_GALV	G21416 gibbon ape
30	1656.5	58.2	522	2	O70651_GALV	O70651 gibbon ape
31	1632.5	57.3	511	1	GAG SMSAV	P03330 simian sarc
32	1472.5	51.7	536	1	GAG_BAEVM	P03341 baboon endo
33	1412.5	49.6	536	2	Q7ZJT8_9GAMR	Q7zjt8 amphotropic
34	1412.5	49.6	1736	2	Q7ZJT6_9GAMR	Q7zjt6 amphotropic
35	1411.5	49.6	535	1	GAG_MLVCB	P27460 cas-br-e mu
36	1409.5	49.5	536	2	O6YTY0_MOUSE	O6yti0 mus musculus
37	1398.5	49.1	534	2	O9YKA0_9GAMR	O9yka0 murine leuk
38	1396.5	49.0	534	2	O6ORS7_9GAMR	O6ofs7 murine leuk
39	1393.5	48.9	1733	2	Q9E7M1_9GAMR	Q9e7m1 dg-75 murin
40	1388	48.7	538	1	GAG_MLVFS	P26807 friend muri
41	1387.5	48.7	538	2	O67457_MLVPR	O67457 friend muri
42	1385.5	48.6	538	2	O12373_9GAMR	O12373 murine leuk
43	1384.5	48.6	521	2	O7ZL05_9GAMR	O7z105 recombinant
44	1384.5	48.6	521	2	O7ZL07_9GAMR	O7z107 recombinant
45	1384.5	48.6	538	2	O41249_MLVRA	O41249 rauscher mu

#### ALIGNMENTS

RESULT 1  
O62706\_PIG  
ID O62706\_PIG PRELIMINARY; PRT; 524 AA.  
AC O62706;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag protein. [pig].  
OS Sus scrofa. [pig].  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
OC Sus.  
OX NCBI\_TaxID=9823;  
RN [1]

RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=miniatore swine;  
RX MEDLINE=98216827; PubMed=9557749;  
RA Akiyoshi D.E., Denaro M., Zhu H., Greenstein J.L., Banerjee P.,  
RA Fishman J.A.;  
RT "Identification of a full-length cDNA for an endogenous retrovirus of  
miniature swine.";  
RL J. Virol. 72:4503-4507(1998).  
DR HSP; AF038600; AAC16763.1; -; mRNA.  
DR HSP; O62706; 7-98.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR003036; Gag\_p30.  
DR InterPro; IPR000840; G\_retro.matrix.  
DR Pfam; PF01140; Gag\_MA; 1.  
DR Pfam; PF02093; Gag\_p30; 1.  
DR Pfam; PF02098; zf-CCHC; 1.  
DR SMART; SM00343; Znf\_C2HC; 1.  
DR PROSITE; PS00158; ZF\_CCHC; 1.  
SQ SEQUENCE 524 AA; 59548 MW; C2D67640A70898B8 CRC64;

Alignment Scores:

Pred. No.:	1.02e-168	Length:	524
Score:	2817.00	Matches:	524
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	98.9%	Indels:	0
DB:	2	Gaps:	0
US-10-723-552-3_COPY_585_2156 (1-1572) x 062706_PIG (1-524)			
Qy	1	ATGGGACAGCGGTGACACACCCCTCTTAGTTGACTCTCGACCATTGAGCTGAAGTTAAA	60
Db	1	MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluVallys	20
Qy	61	TCCAGGGCTCATAAATTTCTGACTGAGTTCAGGTTAAGAGGGACCTTGCGACACTTTCTGTCTC	120
Db	21	SerArgAlaHisAsnLeuSerValGlnVallysGlyProTrpGlnThrPheCysVal	40
Qy	121	TCTGAATGCCGACATTCGATGTTGGATGGCCATCAGAGGGACCTTTAAATTCGAGATT	180
Db	41	SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluile	60
Qy	181	ATCTGGCTGTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG	240
Db	61	IleLeuAlaValLysAlaValIlePheGlnThrGlyProGlySerHisProAspGlnGlu	80
Qy	241	CCCTATATCCTTACGTGGCAAGATTGGCAGAGGATCTCCGCCATGGTGTAAACCATGG	300
Db	81	ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpTrpValysProTrp	100
Qy	301	CTGAATAAGCCAAGAACCCAGGTCCTCCGAAATTCGGCTCTTGGAGAGAAAACAAACAC	360
Db	101	LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis	120
Qy	361	TCGGCTGAAAAGTCAGCCCTCTCTCATATCTATCTACCCCGAGATTGAGGACCCCGCT	420
Db	121	SerAlaGluLysValLysProSerProHisIleTyrProGluIleGluGlnProProAla	140
Qy	421	TGGCCGGAACCCCATCTGTTCCCCACCCCTTATCTGGCACAGGGTCCCGCAGGGGA	480
Db	141	TrpProGluProGlnSerValProProProProTyrLeuAlaGlnGlyAlaAlaArgGly	160
Qy	481	CCCTTTGGCCCTCTCTGGAGCTCCGGCGGTGGAGGACCTGTGTCAGGACTCGAGCCGG	540
Db	161	ProPheAlaProProGlyAlaProAlaValGluGlyProAlaAlaGlyThrArgSerArg	180
Qy	541	AGGGCGCCACCCGGAGCGGACAGACGAGATCGCGACATTACCGCTCGCGACGTACGGC	600
Db	181	ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyrGly	200
Qy	601	CCTCCACACCGGGGGCCAAATTGACGCCCTCCAGTATTGGCCCTTTCTCTGCGAGT	660
Db	201	ProProThrProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp	220
Qy	661	CTCTATAATGGAAAACATAACATCCCTTTCTCGAGAGTCCCCACCGCTCACGGGG	720
Db	221	LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly	240
Qy	721	TTGGTGGAGTCCCTTATGTTCTCTCACCGCTACTTGGGATGATGTCAACAGCTGCTG	780
Db	241	LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspCysGlnGlnLeuLeu	260
Qy	781	CAGACACTTTCACACACGAGGAGGAGAGAGATTCTATTAGAGCTAGAAAATGTT	840
Db	261	GlnThrLeuPheThrThrGluGluArgGluIleLeuLeuGluAlaArgLysAsnVal	280
Qy	841	CCTGGGCGGACGGCGGACCCACGGGTGCAAAATGAGATTGACATGGGATTTCCCTTA	900
Db	281	ProGlyAlaAspGlyArgProThrArgLeuGlnAsnGluIleAspMetGlyPheProLeu	300
Qy	901	ACTGCCCGGGTGGGACTACAAACAGCGCTGAAGGTAGGAGCTTGAAATCTATCGC	960
Db	301	ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg	320

Qy	961	CAGGCTCTGCTGGCGGGTCTCCGGGGCGCTCAAGACGGCCCACTAAATTTGGCTAAGGTA	1020
Db	321	GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal	340
Qy	1021	AGAGAAGTGATGACGGACCCGAATGAACCCCTCTGTTTCTTTTCTTGAGAGCTCTTGAA	1080
Db	341	ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuLeuGlu	360
Qy	1081	GCCTTCAGGCGGTACACCCCTTTTGATCCCACTCAGAGGCCCAAAAGCCCTCAGTGGCT	1140
Db	361	AlaPheArgArgTyrThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla	380
Qy	1141	TTGGCCCTTTATAGACACAGCTCAGCCCTTGATATTAGAAAGAAAGCTTCAGAGACTGGAAGG	1200
Db	381	LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly	400
Qy	1201	TTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAGGACAGAAAGTATATACAAAAGG	1260
Db	401	LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg	420
Qy	1261	GAGACAGAGAGAAAGGACAAAGAAAGAGAGAGAAAGAGAGAAAGGAGGAGAAAGA	1320
Db	421	GluThrGluGluGluArgGlnArgLysGluArgGluArgGluGluArgGluGluArg	440
Qy	1321	CGTAATAAACCGCAAGAGAGAATTTGACTAAGATCTTGGCTGCGAGTGGTTGAAAGGAAA	1380
Db	441	ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys	460
Qy	1381	AGCAATACGGAAGAGAGAGAGATTTTAGAAAAATTAGGTCAAGCCCTAGACAGTCAGGG	1440
Db	461	SerAsnThrGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly	480
Qy	1441	AACCTGGCCATAGACCCCACTCGACAAAGACCAATGTGCATATTGTAAAGAAAGAGA	1500
Db	481	AsnLeuGlyAsnArgThrProLeuAspLysGlnCysAlaTyrCysLysGluArgGly	500
Qy	1501	CAC'TGGGCAAGAACTGCCCCCAAGAGGAAACAAAGACCAAGGATCCTAGCTCTAGAA	1560
Db	501	HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProArgIleLeuAlaLeuGlu	520
Qy	1561	GAAGATAAAGAT 1572	
Db	521	GluAspLysAsp 524	
RESULT 2			
ID	062704_PIG	PRELIMINARY;	PRT; 524 AA.
AC	062704;		
DT	01-AUG-1998	(TrEMBLrel. 07, Created)	
DT	01-AUG-1998	(TrEMBLrel. 07, Last sequence update)	
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)	
DE	Gag protein.		
OS	Sus scrofa (Pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;		
Sus			
NCBI_TaxID=9823;			
[1]			
RP	NEOLEOTIDE SEQUENCE.		
RX	MEDLINE=98216827; PubMed=9557749;		
RA	Akiyoshi D.E., Denaro M., Zhu H., Greenstein J.L., Banerjee P.,		
RA	Fishman J.A.;		
RT	"Identification of a full-length cDNA for an endogenous retrovirus of		
RL	J. Virol. 72:4503-4507(1998).		
DR	EMBL; AF038599; AAC16761.1; -, mRNA.		
DR	HSSP; P03332; 1A6B.		
DR	SMR; 062704; 7-98.		
DR	GO; GO:0019028; C:viral capsid; IEA.		
DR	GO; GO:0003676; F:nucleic acid binding; IEA.		
DR	GO; GO:0005198; F:structural molecule activity; IEA.		
DR	GO; GO:0019068; P:viral assembly; IEA.		
DR	InterPro; IPR003036; Gag_p30.		

DR InterPro; IPR000840; G\_retro\_matrix.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF01140; Gag\_MA; 1.  
 DR Pfam; PF02093; Gag\_p30; 1.  
 DR Pfam; PF00098; zf-CCHC; 1.  
 DR SMART; SM00343; Znf\_C2HC; 1.  
 DR PROSITE; PS00158; ZF\_CCHC; 1.  
 SQ SEQUENCE 524 AA; 59578 MW; 6B54F545098EA1DF CRC64;

Alignment Scores:  
 Pred. No.: 1,82e-168 Length: 524  
 Score: 2813.00 Matches: 522  
 Percent Similarity: 100.0% Conservative: 2  
 Best Local Similarity: 99.6% Mismatches: 0  
 Query Match: 98.8% Indels: 0  
 DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x 062704\_PIG (1-524)

QY 1 ATGGGACAGCGGTGACACCCCTCTTAGTTGACTCTCGACCAATTGGACTGAAGTTAAA 60  
 DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTyrThrGluValys 20  
 QY 61 TCCAGGGCTCATATTTCTCAGTTTCAGTTTAAAGAGGACCTTGGCAGACCTTCTGTGTC 120  
 DB 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTyrGlnThrPheCysVal 40  
 QY 121 TCTGAATGGCGGCACATTCGATGTGGATGGCCATCAGAGGGGACCTTTAAATTCGAGATT 180  
 DB 41 SerGluTyrProThrPheAspValGlyTyrProSerGluGlyThrPheAsnSerGluLeu 60  
 QY 181 ATCTGGCTGTTAAAGCAGTTATTTTCAGACTGACCCGCTCTCATCCGATCAGGAG 240  
 DB 61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
 QY 241 CCTTATATCTTAGTGCAGATTGGCAGAGATCCTCGCCATCGGTAAACCATGG 300  
 DB 81 ProTyrIleLeuThrTyrGlnAspLeuAlaGluAspProProTyrProTyrValLysProTyr 100  
 QY 301 CTGAATAAGCCAGAAAGCCAGGTCCCGGAATTCCTGGCTCTGGAGAGAGAAAAACAACAC 360  
 DB 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
 QY 361 TCGGCTGAAAAGTCAAGCCCTCTCCTCATATCTACCCCGAGATTGAGAGCCACCGGCT 420  
 DB 121 SerAlaGluLysValLysProSerProHisIleTyrProGluIleGluGluProProAla 140  
 QY 421 TGGCCGGAACCCCAATCTGTTCGCCACCCCTTATCTGGCACAGGTCGCCGAGGGGA 480  
 DB 141 TrpProGluProGlnSerValProProProTyrIleLeuAlaGlnGlyAlaAlaArgGly 160  
 QY 481 CCTTGTGCCCTCTCTGGAGCTCCGCGGTGGAGGACCTGCTCGCAGGACTCGGAGCCGG 540  
 DB 161 ProPheAlaProProGlyAlaProAlaValGluGlyProSerAlaGlyThrArgSerArg 180  
 QY 541 AGGGCGCCACCCCGGAGCGGACAGACGAGATTCGCGACATTACCGCTCGCGACCTACGGC 600  
 DB 181 ArgGlyAlaThrProGluAlaThrAspGluIleAlaThrLeuProLeuAlaThrTyrGly 200  
 QY 601 CTCTCCACACGGGGGCGCAATTGACGCCCTTCAGATATTGGCCCTTTCTTCTGCGAGAT 660  
 DB 201 ProProThrProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
 QY 661 CTCTATATATGGAACCTAACCATCTTCTCGGAGGATCCCAACGCCCTCACCGGGG 720  
 DB 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
 QY 721 TTGGTGGAGTCCCTTATGTTCTCTCACGAGCTACTTGGGATGATTGTCAACAGCTCTG 780  
 DB 241 LeuValGluSerLeuMetPheSerHisGlnProThrTyrPheAspCysGlnGlnLeu 260  
 QY 781 CAGACACTCTTCAACCCGAGGCGGAGAGAAATTCATTAGAGGCTAGAAAAATGTT 840

Db 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280  
 QY 841 CCTGGGCGGACGCGGCGGACCCACCGGTTGCAAAATGAGATTGACATCGGATTTCCCTTA 900  
 Db 281 ProGlyAlaAspGlyArgProThrArgLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
 QY 901 ACTCGCCCGGTTGGGACTCAACACCGCTCAAGGTAGGGAGAGCTTGAAATCTATCGC 960  
 Db 301 ThrArgProGlyTyrAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
 QY 961 CAGGCTCTGGTGGGGTCTCCGGGGCCCTCAAGACGGCCCACTAATTTGGCTAAGGTA 1020  
 Db 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgProThrAsnLeuAlaLysVal 340  
 QY 1021 AGAGAAGTGTAGCAGGACCGAATGAACCCCTCTGTTTCTTTTCTTGAGAGGCTCTTGAA 1080  
 Db 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuLeuGlu 360  
 QY 1081 GCCTTCAGGCGGTACACCCCTTTTGTATCCCACTCAGAGGCCCAAAAGCCCTCAGTGCT 1140  
 Db 361 AlaPheArgTyrThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
 QY 1141 TTGCGCTTATAGACAGTCTGAGCTTCGATATTAGAAAGAGCTTCAGAGACTGGAAGG 1200  
 Db 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGly 400  
 QY 1201 TTACAGGAGGCTGATTTACGTGATCTAGTGAAGGAGCAGAGAAAGTATATTACAAAAG 1260  
 Db 401 LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg 420  
 QY 1261 GAGACAGAAGAAGAGGGAACAAAGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 Db 421 GluThrGluGluArgGluGlnArgLysGluArgGluArgGluArgGluArgGluArg 440  
 QY 1321 CGTAATAACCGGCAAGAGAAATTTGACTAAGATCTTGGCTGAGTGGTTGAAGGGAAA 1380  
 Db 441 ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaValGluGlyLys 460  
 QY 1381 AGCAATACGAAAGAGAGAGAGATTTTAGGAAAATATTAGTTCAGGCCCTAGACAGTCAGG 1440  
 Db 461 SerAsnThrGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
 QY 1441 AACCTGGCAATAGGACCCCACTCGACAAGCAACATGTGCATATTGTAAAGAAAGAGGA 1500  
 Db 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluArgGly 500  
 QY 1501 CACTGGCAGGAACTCCGCCCAAGAGGGAACAAAGGACCAAGGATCCTAGCTCTAGTA 1560  
 Db 501 HisTyrAlaArgAsnCysProLysLysGlyAsnLysGlyProArgIleLeuAlaLeuGlu 520  
 QY 1561 GAAGATAAAGAT 1572  
 Db 521 GluAspLysAsp 524

RESULT 3  
 Q5OGQ9\_9GAMR  
 ID Q5OGQ9\_9GAMR PRELIMINARY; PRT; 524 AA.  
 AC Q5OGQ9\_9GAMR  
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
 DE Gag protein.  
 OS Porcine endogenous retrovirus C/A.  
 OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
 OC Gamaretrovirus; 1-Mammalian type C virus group.  
 NCBI\_TaxID=286072;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=A14/220;  
 RX PubMed=15564496; DOI=10.1128/JVI.78.24.13880-13890.2004;  
 RA Bartosch B., Stefanidis D., Myers R., Weiss R., Patience C.,  
 RA Takeuchi Y.;  
 RT "Evidence and consequence of porcine endogenous retrovirus

```
RT recombination.";
RL J. Virol. 78:13880-13890 (2004).
DR EMBL; AY570980; AAT77166.1; -, Genomic_DNA.
DR SMR; Q5QG09; 7-98.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0019068; P:viral assembly; IEA.
DR InterPro; IPR003036; Gag_p30.
DR InterPro; IPR000840; G_retro_matrix.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01140; Gag_MA; 1.
DR Pfam; PF02093; Gag_p30; 1.
DR Pfam; PF00098; ZF-CCHC; 1.
DR SMART; SM00343; ZNF_C2HC; 1.
DR PROSITE; PS0158; ZF_CCHC; 1.
SQ SEQUENCE 524 AA; 59534 MW; 24805E0EBAA46E1F CRC64;

Alignment Scores:
Pred. No.: 2,1e-168 Length: 524
Score: 2812.00 Matches: 522
Percent Similarity: 100.0% Conservative: 2
Best Local Similarity: 99.6% Mismatches: 0
Query Match: 98.7% Indels: 0
DB: 2 Gaps: 0

US-10-723-552-3_COPY_585_2156 (1-1572) x Q5QG09_9GAMR (1-524)
QY 1 ATGGACAGAGCGGTGACGACCCCTCTTAGTTTGACTCTCGACCAATTGGACTGAAGTTAAA 60
DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisThrThrGluVallys 20
QY 61 TCCAGGCTCATATTTGTCAGTTAGTTCAGTTAGAGGACCTTGGCAGACTTCTGTGTC 120
DB 21 SerArgAlaHisAsnLeuSerValGlnVallysGlyProThrGlnThrPheCysVal 40
QY 121 TCTGAATGCGCGACATTCGATGTTGCGATCGGCATCAGAGGGGACCTTTAAATTCGAGATT 180
DB 41 SerGlnThrProThrPheAspValGlyThrProSerGluGlyThrPheAsnSerGluIle 60
QY 181 ATCTGGCTGTTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 240
DB 61 IleLeuAlaValLysAlaIleIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80
QY 241 CCTATATCTTACCTGGCAGAGATTGGCAGAGATCTCCGCCCTGGGTTAAACCATGG 300
DB 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProProThrValLysProThr 100
QY 301 CTGAATAGCCAAAGAACCCAGGTCCTCCGAAATCTGGCTCTTGGAGAGAAAAACAACAC 360
DB 101 LeuAsnLysProArgLysProGlyProGlyProGlyIleLeuAlaLeuGlyGluLysAsnLysHis 120
QY 361 TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGACCCCGGCT 420
DB 121 SerAlaGluLysValLysProSerProHisIleTyrProGluIleGluGluProAla 140
QY 421 TGGCCGGAACCCCAATCTGTTCCCCACCCCTTATCTGGACACGGGTCCCGCGAGGGA 480
DB 141 TrpProGluProGlnSerValProProProTyrLeuAlaGlnGlyAlaAlaArgGly 160
QY 481 CCCTTTGCCCTCTGAGAGCTCCGCGGTGGAGGACCTGTGACAGGAGCTCGGAGCCGG 540
DB 161 ProPheAlaProProGlyAlaProAlaValGluGlyProAlaAlaGlyThrArgSerArg 180
QY 541 AGGGCGGCCACCCCGAGCGGACAGACGAGATCGGCACATTACCGCTCGCACGTCACGGC 600
DB 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyrGly 200
QY 601 CTCCACACCGGGGGGCAATTGAGCCCTCCAGTATTGGCCCTTCTCTTCGACAT 660
DB 201 ProProThrProGlyGlyGlnLeuGlnProLeuGlnTyrThrProPheSerSerAlaAsp 220
QY 661 CTCATATAATTGGAAACTAAACCATCCCTCTTCGGAGGATCCCAACAGCGCTCACGGGG 720
```

```
Db 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240
QY 721 TTGGTGGAGTCCCTTAATGTTCTCTCACAGCTACTTGGGATGATTGTCAACAGCTGCTG 780
DB 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
QY 781 CAGACACTCTTCACAACCGAGGACGAGAGAGAAATCTATTAGAGGCTTAGAAAAAATGTT 840
DB 261 GlnThrLeuPheThrThrGluGluArgGluIleLeuLeuGluAlaAaArgLysAsnVal 280
QY 841 CCTGGGCGCCGACGCGGACCCACCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA 900
DB 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300
QY 901 ACTGCCCCGGTGGGACTACACACCGCTGAAGGTAGGAGAGCTTGAAATCTATCGC 960
DB 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320
QY 961 CAGGCTCTGTGGCGGCTCTCCGGGCGCCTCAAGACGGCCCACTAAATTTGCTAAGGTA 1020
DB 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgProThrAsnLeuAlaLysVal 340
QY 1021 AGAGAAGTGTATGTCAGGACCCGAATGAACCCCTCTCTGTTTTTCTTGAGAGGCTCTTGAA 1080
DB 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuLeuGlu 360
QY 1081 GCCTTCAGGCGGTACACCCCTTTGATCCACCTCAGAGGCCCAAAAGCCTCAGTGGCT 1140
DB 361 AlaPheArgArgTyrThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380
QY 1141 TTGSCCTTTATAGACAGCTCAGCCTTGATATTAAGAAAGACTTCAGAGACTTGAAGGG 1200
DB 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400
QY 1201 TTACAGGAGGCTGAGTTACGTGATCTAGTAGGAGGACAGAAAGTATATTACAAAAGG 1260
DB 401 LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg 420
QY 1261 GAGACAGAGAAAGGAAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 421 GluThrGluGluGluArgGluGlnArgLysGluArgGluArgGluArgGluArgGluArg 440
QY 1321 CGTAATAACCGCAGAGAGAAATTTGACTAAGATCTTGGCTGCGAGTGGTTGAAGGGAAA 1380
DB 441 ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValGluGlyLys 460
QY 1381 AGCAATACGGAAGAGAGAGAGATTTTAGGAAAAATTAGGTACGCCCTTAGACAGTCAGG 1440
DB 461 SerAsnThrGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480
QY 1441 AACCTGGGCAATAGAGCCCACTCGACAGGACCAATGTGCATATTGTAAAGAAAGAGA 1500
DB 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluArgGly 500
QY 1501 CACTGGGCAAGAACTGCCCCCAAGAGGAAACAAAGGACCAAGGATCCTAGCTCTAGAA 1560
DB 501 HisTrpAlaArgAsnCysProLysGlyAsnLysGlyProArgIleLeuAlaLeuGlu 520
QY 1561 GAAGATAAAGAT 1572
DB 521 GluAspLysAsp 524

RESULT 4
Q4VFZ3_9GAMR
ID Q4VFZ3_9GAMR PRELIMINARY; PRT; 524 AA.
AC Q4VFZ3;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Gag protein.
GN Name=gag;
OS Porcine endogenous retrovirus C/A.
```



OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
 OX Gammaretrovirus; 1-Mammalian type C virus group.  
 RN NCBI\_TaxID=286072;  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C/A 4;  
 RA Denner J., Karlas A., Votter J.;  
 RT "Nucleotide sequence and extended characterization of a high passage  
 RL human cell-adapted recombinant PERV-C/A.";  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY53542; AAY28927.1; -; Genomic\_DNA.  
 DR InterPro; IPR000840; G\_retro\_matrix.  
 DR InterPro; IPR003036; Gag\_p30.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF01140; Gag\_P30; 1.  
 DR Pfam; PF02093; Gag\_P30; 1.  
 DR Pfam; PF00098; zf-CCHC; 1.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR SMART; SM00343; Znf\_C2HC; 1.  
 DR PROSITE; PS0158; ZF\_CCHC; 1.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 524 AA; 59449 MW; 3825CCA687E1F450 CRC64;

Alignment Scores:  
 Pred. NO.: 5,14e-166 Length: 524  
 Score: 2774.00 Matches: 515  
 Percent Similarity: 99.4% Conservative: 6  
 Best Local Similarity: 99.3% Mismatches: 3  
 Query Match: 97.4% Indels: 0  
 DB: 2 Gaps: 0

US-10-723-552-3\_copy\_585\_2156 (1-1572) x Q4VPZ3\_9GAMR (1-524)

QY	1	ATGGGACAGAGGGTACGACCCCTCTAGTTTGA	CTCGACCATTCGAGCTGAAGTTAAA	60
DB	1	MettGlyGlnThrValThrProLeuSerLeuThrLeuAspHisPrtThrGluValLys		20
QY	61	TCCAGGCTCATATTTGTCAAGTTTAAAGAGGACCTTCGCGACAGCTTTCTGTGTC		120
DB	21	SerArgAlaHisLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysVal		40
QY	121	TCGTAATGGCGGACATTCGATTTGGATGGCCATCAGAGGGACCTTTAATTCGAGATT		180
DB	41	SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle		60
QY	181	ATCTGCTGCTTAAAGCAGTTATTTTCAGACTGACCGGCTCTCATCCGATCAGGAG		240
DB	61	IleLeuAlaValLysAlaIleIlePheGlnThrGlyProGlySerHisProAsnGlnGlu		80
QY	241	CCCTATATCTTACGTGGCAAGATTTCGACAGAGGATCTCCGCCATCGGTTAAACCATGG		300
DB	81	ProTyrlleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp		100
QY	301	CTGAATTAAGCAAGAAAGCCAGGTCCCGCAATTCCTGGCTCTTGAGAGAGAAAAACAAC		360
DB	101	LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis		120
QY	361	TCGGCTGAAAAGTCAAGCCCTCTCCATATCTACCCCGAGATTGAGGAGCCACCGGCT		420
DB	121	SerAlaGluLysValLysProSerProHisIleIleTrpProGluIleGluProProAla		140
QY	421	TGGCCGGAACCCCAATCTGTTCCCCACCCCTTATCTGGCACAGGTCGCGGAGGGGA		480
DB	141	TrpProGluProGlnSerValProProProProTrpLeuAlaGlnGlyAlaAlaArgGly		160
QY	481	CCCTTTGCCCTCTCTGGAGCTCCGGCGGTGGAGGACCTGCTCGAGGGACTCGAGCCGG		540
DB	161	ProSerAlaProProGlyAlaProAlaValGluGlyProAlaAlaGlyThrArgSerArg		180
QY	541	AGGGGCCACCCCGGAGCGGACGACGAGATTCGCGACATTACCGCTGCGCAGCTACGGC		600
DB	181	ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyArg		200

RESULT 5

Q6UUZ3\_9GAMR

ID Q6UUZ3\_9GAMR PRELIMINARY;

PRT; 524 AA.

AC Q6UUZ3;

QY	601	CTCTCCACACCGGGGGCCAAATTGACGCCCTCCAGATTCAGCCCTTTCTTCTGCGAGAT	660
DB	201	ProProThrProGlyGlyGlnLeuGlnProLeuGlnTrpProPheSerSerAlaAsp	220
QY	661	CTCTATAATTGGAAACTAACCATCCCTCTTCGGAGGATCCCAACGCTCACGGGG	720
DB	221	LeuTyraSerTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly	240
QY	721	TTGGTGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGATGATGTCAACAGCTGTG	780
DB	241	LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu	260
QY	781	CAGACACTTTCACACCCGAGGACGAGAGAAATTCATTAGAGGCTAGAAAAAATGTT	840
DB	261	GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal	280
QY	841	CTTGGGGCCGACGGGCGACCCACCGCGTTCGAAAATGAGATTGATGAGTTTCCCTTA	900
DB	281	ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu	300
QY	901	ACTCGCCCGGTTGGGACTACACACGCTGAAGTAGGAGAGCTTGAAAATCTATCGC	960
DB	301	ThrArgProGlyTrpAspTrpAsnThrAlaGluGlyArgGluSerLeuLysIleTyArg	320
QY	961	CAGGCTCTGTCGGCGGTCTCCGGGCGCTCAAGACGGCCCACTAATTTGGCTTAAGGTA	1020
DB	321	GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal	340
QY	1021	AGAGAGTGATGACAGGACCGGAATGAACCCCTCTGTTTCTTGTAGAGGCTCTTGAA	1080
DB	341	ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluTrpLeuGlu	360
QY	1081	GCCTTCAGGCGGTACACCCCTTTTGATCCACCTTCAGAGGCCCAAGCCTCAGTCGCT	1140
DB	361	AlaPheArgArgTrpProPheAspProThrSerGluAlaGlnLysAlaSerValAla	380
QY	1141	TTGGCTTTTAGGACAGTCTGAGCTTGGATATTAGAAAGAGCTTCAGAGACTCGAAGG	1200
DB	381	LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly	400
QY	1201	TTACAGAGGCTCAGTTACGTGATCTAGTGAAGAGGACAGAGAAAGTATTATTACAAAG	1260
DB	401	LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyTrpLysArg	420
QY	1261	GAGACAGAAACAAGGGAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1320
DB	421	GluThrGluGluGluArgGluGlnArgLysGluArgGluGluArgGluGluArgGluGlu	440
QY	1321	CGTAATAAACCGCAAGAGAGAAATTTGACTTAAGATCTTGGCTGCGTGGTTGAAGGAAA	1380
DB	441	ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValValGluGlyLys	460
QY	1381	AGCAATACGAAAGAGAGAGAGATTTTAGAAAATTTAGGTTCAGGCCCTAGACAGTCAGG	1440
DB	461	SerAsnThrGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly	480
QY	1441	AACTCGGCATATAGGACCCCACTCGACACAGGACCAATGTCATATTGTAAGAAGAGGA	1500
DB	481	AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyCysLysGluLysGly	500
QY	1501	CACCTGGCAGGAATCTGCCCAAGAGGGAACAAAGGACCAAGGATCTTAGCTCTAGAA	1560
DB	501	HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyLeuLysValLeuAlaLeuGlu	520
QY	1561	GAAGATAAAGAT 1572	
DB	521	GluAspLysAsp 524	

DT 05-JUL-2004 (TremBLrel. 27, Created)  
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)  
DT 10-MAY-2005 (TremBLrel. 30, Last annotation update)  
DE Gag protein.  
GN Name=gag;  
OS Porcine endogenous retrovirus.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gammaretrovirus; 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed:14963152; DOI=10.1128/JVI.78.5.2502-2509.2004;  
RA Scobie L., Taylor S., Wood J.C., Suling K.M., Quinn G., Meikle S.,  
RA Patience C., Schuurman H.J., Onions D.E.;  
RT "Absence of replication-competent human-tropic porcine endogenous  
RT retroviruses in the germ line DNA of inbred miniature swine.";  
RL J. Virol. 78:2502-2509(2004).  
DR EMBL; AY368585; AAQ73934.1; -; Genomic DNA.  
DR EMBL; AY368584; AAQ73936.1; -; Genomic DNA.  
DR HSP; P03332; IAGB.  
DR SMR; Q6UJZ3; 7-98.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR003036; Gag\_p30.  
DR InterPro; IPR000840; G\_retro\_matrix.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF01140; Gag\_MA; 1.  
DR Pfam; PF02093; Gag\_p30; 1.  
DR Pfam; PF00098; zf\_CCHC; 1.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR SMART; SM00343; Znf\_C2HC; 1.  
DR PROSITE; PS00158; ZF\_CCHC; 1.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 524 AA; 59474 MW; 72178B9541475920 CRC64;  
  
Alignment Scores:  
Pred. No.: 524  
Score: 2757.00 Matches: 511  
Percent Similarity: 99.0% Conservative: 8  
Best Local Similarity: 97.5% Mismatches: 5  
Query Match: 96.8% Indels: 0  
DB: 2 Gaps: 0  
  
US-10-723-552-3\_COPY\_585\_2156 (1-1572) x Q6UJZ3\_9GAMR (1-524)  
QY 1 ATGGGACAGAGCGTGACGACCCCTCTTAGTTGACTCTCGACCATTTGGACTGAAGTTAAA 60  
DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpTrpGlnThrGluValLys 20  
QY 61 TCCAGGCTCATAAATTGTCAGTTAGAGTAAAGAGGACCTTGGCAGACTTTCTGTGTC 120  
DB 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysVal 40  
QY 121 TCTGAATGGCCGACATTGATGTTGGATGGCCATCAGAGGGACCTTAAATTCGAGATT 180  
DB 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
QY 181 ATCTCGGCTGTTAAGCAGATTATTTTTCAGACTGACCGGCTCTCATCCCGATCAGGAG 240  
DB 61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
QY 241 CCCTATATCTTACCTGGCAAGATTGGCAGAGGATCTCCGCCATGGGTTAAACCATGG 300  
DB 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100  
QY 301 CTGAATAAGCCNAGAACCGAGTCCCGAATCTGGCTCTTGGGAGAGAAAAAACHAACAC 360  
DB 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
QY 361 TCGGCTGAAAAAGTAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGAGCCACCGGCT 420

DB 121 SerAlaGluLysValLysProSerProHisIleTyrProGluIleGluGluProProAla 140  
QY 421 TGGCCGGAAACCCCAATCTGTTCCCAACCCCTTATCTGGCACAGGGTCCCGCAGGGGA 480  
DB 141 TrpProGluProGlnSerValProProProTyrProAlaGlnGlyAlaAlaArgGly 160  
QY 481 CCCTTTCCCTCTCTGGAGCTCCGGCGGTGGAGGACCTGCTGCAGGAGACTCGAGACGG 540  
DB 161 ProSerAlaProProGlyAlaProAlaValGluGlyProAlaAlaGlyThrArgSerArg 180  
QY 541 AGGGCGCCACCCCGAGCGGACAGAGATCCGACATTACCGCTCGCACGCTACGGC 600  
DB 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyrGly 200  
QY 601 CCTCCACACCGGGGGGCAATTCGAGCCCTCAGATATTGGCCCTTTCTTCTGCAGAT 660  
DB 201 ProProIleProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
QY 661 CTCTATAATTGGAACCTAACCATCCCTTTCTCGAGGATCCCAACGCTCACGGGG 720  
DB 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
QY 721 TTGGTGGAGTCCCTTATGTTCTCTCACAGCTACTTGGGATGATTGTCAACAGCTGCTG 780  
DB 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260  
QY 781 CAGACACTCTTCAACACCGAGGAGGAGAGAGAAATCTTATTAGAGGCTAGAAAAATGTT 840  
DB 261 GlnThrLeuPheThrThrGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280  
QY 841 CCTGGGCGGACGCGGACCCACCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA 900  
DB 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGlnIleAspMetGlyPheProLeu 300  
QY 901 ACTCGCCCGGTTGGGACTACAACACGCTGAAGGTAGGAGAGCTTGAAATCTATCGC 960  
DB 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
QY 961 CAGCTCTGGTGGCGGCTCTCCGGGCGCTCAAGACGCGCCCACTAATTTGGCTAAGTA 1020  
DB 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
QY 1021 AGAGAGTGATGACAGGACCGAATGAACCCCTCTGTTTTCTTTGAGAGCTCTTGAA 1080  
DB 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
QY 1081 GCCTTCAGGCGGTACACCCCTTTTATGCCCACTCAGAGGCCCAAAAAGCCTCAGTGCT 1140  
DB 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
QY 1141 TTGGCCTTTATAGACAGTCAAGCTTGGATATTAGAAGAAGCTTCAGAGACTGGAAGG 1200  
DB 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
QY 1201 TTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGAGGACAGAGAAAGTATATTACAAAAG 1260  
DB 401 LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg 420  
QY 1261 GAGACAGAAAGAAAGGAAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 421 GluThrGluGluGluArgGluGlnArgLysGluArgGluArgGluGluArgGluArg 440  
QY 1321 CGTAAATAACCGCAGAGAGAGAAATTTGACTAAGATCTTGGCTGCGAGTGGTGAAGGAAA 1380  
DB 441 ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValValGluGlyLys 460  
QY 1381 AGCAATACGGAAGAGAGAGAGATTTTAGGAAAATTTAGGTCCAGGCCCTTAGACAGTCAGG 1440  
DB 461 SerAsnArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
QY 1441 AACCTGGGCAATAGGACCCCACTCGACAAAGACCAATGTGCATATTGTAAAGAAAGAGA 1500

Db	481	AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly	500
Qy	1501	CACTGGGCAAGAACTGCCCAAGGAAACAAAGACCAAGGATCCTAGCTTAGAA	1560
Db	501	HisTrpAlaArgAspCysProLysLysGlyAsnLysGlyLeuLysValLeuAlaLeuGlu	520
Qy	1561	GAAGATAAAGAT	1572
Db	521	GluAspLysAsp	524
RESULT 6			
ID	Q6T3V4_9GAMR	PRELIMINARY;	PRT; 524 AA.
AC	Q6T3V4;		
DT	05-JUL-2004	(TrEMBLrel. 27, Created)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)	
DE	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)	
DE	Gag protein.		
GN	Name=gag;		
OS	Porcine endogenous retrovirus.		
OC	Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;		
OC	Gammaretrovirus; 1-Mammalian type C virus group.		
OX	NCBI_TaxID=61673;		
EN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Scobie L., Taylor S., Wood J.C., Suling K.M., Quinn G., Patience C.,		
RA	Schuurman H.-J., Onions D.E.;		
RT	"Absence of replication competent human-tropic porcine endogenous		
RT	retroviruses in the germ line DNA of inbred miniature swine.";		
RL	J. Virol. 78:0-0(2004).		
DR	EMBL; AY437841; AAR28102.1; -; Genomic_DNA.		
DR	HSSP; P03332; 1A6B.		
DR	SMR; Q6T3V4; 7-98.		
DR	GO; GO:0019028; C:viral capsid; IEA.		
DR	GO; GO:0046872; F:metal ion binding; IEA.		
DR	GO; GO:0003676; F:nucleic acid binding; IEA.		
DR	GO; GO:0005198; F:structural molecule activity; IEA.		
DR	GO; GO:0019068; P:viral assembly; IEA.		
DR	InterPro; IPR003036; Gag_P30.		
DR	InterPro; IPR000840; G_retro_matrix.		
DR	InterPro; IPR001878; Znf_CCHC.		
DR	Pfam; PF01140; Gag_MA; 1.		
DR	Pfam; PF02093; Gag_P30; 1.		
DR	Pfam; PF00098; zf_CCHC; 1.		
DR	PRINTS; PR00939; C2HCZNFINGER.		
DR	SMART; SM00343; Znf_C2HC; 1.		
DR	PROSITE; PS00158; zf_CCHC; 1.		
KW	Metal-binding; Zinc; Zinc-finger.		
SQ	SEQUENCE 524 AA; 59516 MW; 4B6B6C68FE930E32 CRC64;		
Alignment Scores:			
Pred. No.:	2,56e-164	Length:	524
Score:	2747.00	Matches:	509
Percent Similarity:	98.7%	Conservative:	8
Best Local Similarity:	97.1%	Mismatches:	7
Query Match:	96.5%	Indels:	0
DB:	2	Gaps:	0
US-10-723-552-3_COPY_585_2156 (1-1572) x Q6T3V4_9GAMR (1-524)			
Qy	1	ATGGGACAGCGGTGAGACCCCTCTAGTTGTGACTCTCGACCAATTCGACTGAAGTTAAA	60
Db	1	MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValLys	20
Qy	61	TCCAGGCTCATATTTGTCAGTTTCAGTTTAAGGAGGACCTTGGCAGACTTCTGTGTC	120
Db	21	SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysVal	40
Qy	121	TCTGAATGGCGGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCGAGATT	180
Db	41	SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle	60
Qy	181	ATCCTGGCTGTAAAGCAGTATTATTTTCAGACTGGACCCGGCTCTCATCCGATCAGGAG	240

Db	61	IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu	80
Qy	241	CCCTATATCTTACGTGCGAAGATTTCGCAGAGGATCTCCCGCATCGGTTTAAACCATGG	300
Db	81	ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp	100
Qy	301	CTGAATAAGCCACAGAAAGCCAGGTCCCGCAATTCCTGGCTCTTGGAGAGAAAAACAAACAC	360
Db	101	LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis	120
Qy	361	TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGAGCCACCGCT	420
Db	121	SerAlaGluLysValLysProSerProHisIleTyrProGluIleGluGluProProAla	140
Qy	421	TGGCCGGAACCCCAATCTGTTCCCGCCACCCCTTATCTGGCAGAGGTGCCCGAGGGA	480
Db	141	TrpProGluProGlnSerValProProProTyrProAlaGlnGlyAlaAlaArgGly	160
Qy	481	CCCTTTGCCCTCTCTGGAGCTCCCGCGGTGGAGGACCTGTCTGAGGAGCTCGGAGCCGG	540
Db	161	ProSerAlaProProGlyAlaProAlaValGluGlyProAlaAlaGlyThrArgSerArg	180
Qy	541	AGGGCGCCACCCCGAGCGGACAGAGAGATCCGCACTTACCGCTGCGCACGCTACGCG	600
Db	181	ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyrGly	200
Qy	601	CCTCCACACCGGGGGCCAAATTGCAGCCCTCCAGTATTGGCCCTTTCTTCTCGAGAT	660
Db	201	ProProIleProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp	220
Qy	661	CTCTATAATTGGAAAAATAACCATCCCGCTTCTCGGAGGATCCCAACGCTCACCGGG	720
Db	221	LeuTyrAsnTrpLysThrAsnHisProPheSerGluAspProGlnArgLeuThrGly	240
Qy	721	TTGGTGAGTCCCTTATGTTCTCTCACAGCTACTCTGGAGATGATTGTCAACAGCTGCTG	780
Db	241	LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu	260
Qy	781	CAGACACTCTTCAACCGAGGAGCAGAGAGAACTTCTATTAGAGGCTAGAAAAAATGTT	840
Db	261	GlnThrLeuPheThrThrGluGluArgGluGluIleLeuLeuGluAlaArgLysAsnVal	280
Qy	841	CCTGGGCGCCAGCGGCGACCCAGCGGTTGCAAAATCAGATTGACATGGGATTTCCCTTA	900
Db	281	ProGlyThrAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu	300
Qy	901	ACTGCGCCCGTTGGGACTACAAACACCGCTGAAGGTAGGAGAGCTTGAATAATCTATCG	960
Db	301	IleArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg	320
Qy	961	CAGGCTCTGCTGGGGGTCTCCGGGGCGCTCAAGAGCGGCCACTAATTTGGCTAAGGTA	1020
Db	321	GlnAlaLeuValAlaGlyLeuArgLysValSerArgArgProThrAsnLeuAlaLysVal	340
Qy	1021	AGAGAACTGATCAGGAGCAGCAATGAACCCCTCTCTTTCTTTGAGAGGCTCTTGAA	1080
Db	341	ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu	360
Qy	1081	GCCTTCAGGCGGTACCCCTTTTGTATCCACCTCAGAGGCCCAAAAGCCTCAGTGCT	1140
Db	361	AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla	380
Qy	1141	TTGCGCTTTATAGCAGACTCAGCTCGATATTAGAAAGAGCTTCAGAGACTTGGAGGG	1200
Db	381	LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly	400
Qy	1201	TTACAGGAGCTCAGTTACGTGATCTAGTGAAGGAGGACAGAAAGTATATTACAAAAG	1260
Db	401	LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg	420
Qy	1261	GAGACAGAAAGAAAGGGAACAAAGAAAAAGAGAGAGAAAGAGAGGAAAGGAGGAAGA	1320

```
Db      421  GluThrGluGluGluArgGluGlnArgLysGluArgGluArgGluGluArgGluGluArg 440
Qy      1321 CGTAAATAACGGCAAGAGAAGAAATTTGACTAAGATCTTGGCTGCGAGTGGTTGAAGGGAAA 1380
Db      441  ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys 460
Qy      1381  AGCAATACGGAAGAGAGAGAGATTTTAGGAAATTTAGTCCAGGCCCTAGACAGTCAGGG 1440
Db      461  SerAsnArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480
Qy      1441  AACCTGGGCAATAGGACCCCACTCGACAAGGACCAATGTGCATATTTGTAAGAAGAGAGA 1500
Db      481  AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500
Qy      1501  CACTGGGCAAGAACTGCCCCAAGAGGGAACAAAGGACCAAGATCCTAGCTCTAGAA 1560
Db      501  HisTrpAlaArgAspCysProLysLysGlyAsnLysGlyLeuLysValLeuAlaLeuGlu 520
Qy      1561  GAAGATAAAGAT 1572
Db      521  GluAspLysAsp 524
```

## RESULT 7

```
ID      Q8UMA0_9GAMR PRELIMINARY; PRT; 524 AA.
AC      Q8UMA0;
DT      01-MAR-2002 (TRENBLrel. 20, Created)
DT      01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT      01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE      Group specific antigen.
GN      Name=gag;
OS      Porcine endogenous retrovirus.
OC      Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC      Gammatetovirus; 1-Mammalian type C virus group.
OX      NCBI_TaxID=61673;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RX      MEDLINE=20219394; PubMed=10756014;
RX      DOI=10.1128/JVI.74.9.4028-4038.2000;
RA      Czauderna F., Fischer N., Boller K., Kurth R., Toenjes R.R.;
RT      "Establishment and characterization of molecular clones of porcine
RL      J. Virol. 74:4028-4038(2000).
RN      [2]
RP      NUCLEOTIDE SEQUENCE.
RX      MEDLINE=20219394; PubMed=11861838;
RX      DOI=10.1128/JVI.76.6.2714-2720.2002;
RA      Niebert M., Rogel-Gaillard C., Chardon P., Toenjes R.R.;
RT      "Characterization of Chromosomally Assigned Replication-Competent Gamma
RT      Porcine Endogenous Retroviruses Derived from a Large White Pig and
RT      Expression in Human Cells.";
RL      J. Virol. 76:2714-2720(2002).
DR      EMBL; AJ279056; CAC82501.1; -; Genomic_DNA.
DR      HSP; P03332; 1A6B.
DR      SMR; Q8UMA0; 7-98.
DR      GO; GO:0019028; C:viral capsid; IEA.
DR      GO; GO:0046872; C:metal ion binding; IEA.
DR      GO; GO:0003676; F:nucleic acid binding; IEA.
DR      GO; GO:0005198; F:structural molecule activity; IEA.
DR      GO; GO:0019068; P:viral assembly; IEA.
DR      InterPro; IPR003036; Gag_p30.
DR      InterPro; IPR000840; G_retro_matrix.
DR      InterPro; IPR001878; Znf_CCHC.
DR      Pfam; PF01140; Gag_MA; 1.
DR      Pfam; PF02093; Gag_p30; 1.
DR      Pfam; PF00098; zf_CCHC; 1.
DR      PRINTS; PR00939; C2HCZNFINGER.
DR      SMART; SM00343; ZNF_C2HC; 1.
DR      PROSITE; PS50158; ZF_CCHC; 1.
DR      Metal-binding; Zinc; Zinc-finger.
SQ      SEQUENCE 524 AA; 59502 MW; A5PD9F31FB3291AE CRC64;
```

Alignment Scores:

```
Pred. No.: 2.56e-164 Length: 524
Score: 2747.00 Matches: 508
Percent Similarity: 98.9% Conservative: 10
Best Local Similarity: 96.9% Mismatches: 6
Query Match: 96.5% Indels: 0
DB: 2 Gaps: 0

US-10-723-552-3_copy_585_2156 (1-1572) x Q8UMA0_9GAMR (1-524)
Qy      1  ATGGGACAGAGCGTGAGACGCCCTCTTAGTTTGTACTCTCGACCAATTGGACTGAAGTTAAA 60
Db      1  MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValLys 20
Qy      61  TCCAGGGCTCATATTTGTCTAGTTTCAGTTTCAAGTAAAGAGGACCTTCGCGACATTTCTGTCTC 120
Db      21  SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysVal 40
Qy      121  TCTGAATGGCGGACATTCGATTTGGATGGCCATCAGAGGGGACCTTTAAATTCGAGATT 180
Db      41  SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60
Qy      181  ATCTGCTGTTTAAAGCAGATTATTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 240
Db      61  IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80
Qy      241  CCTATATCTTACGTGCGCAAGATTTCGCGAGGATCTCCGCCATCGGTTTAAACCATGG 300
Db      81  ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100
Qy      301  CTGAATAAGCCAAAGAACCCAGGTCCCGAATTCGCTCTTGGAGAGAAAAACAAACAC 360
Db      101  LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120
Qy      361  TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGAGCCACCGGCT 420
Db      121  SerAlaGluLysValLysProSerProHisIleTyrProGluIleGluGluProProAla 140
Qy      421  TGGCGGGAACCCCAATCTGTTCGCCCCACCCCTTATCTGGCCACAGGGTCCGCGAGGGGA 480
Db      141  TrpProGluProGlnSerValProProProTrpProProTrpProAlaGlnGlyAlaAlaArgGly 160
Qy      481  CCCTTGCCCCCTCTCTGGAGCTCCGGCGGTGAGGAGACCTCTGTCAGGAGACTCGGAGCGG 540
Db      161  ProSerAlaProProGlyAlaProAlaValGluGlyProValAlaGlyThrArgThrArg 180
Qy      541  AGGGGCGCCACCCCGGAGCGGACAGAGATCCGGACATTACCGCTCCGCGACGTACGGC 600
Db      181  ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyrGly 200
Qy      601  CCTCCACACCGGGGGCCCAATTCAGCCCTCCAGTATTCGCCCTTTTCTTCTGCAGAT 660
Db      201  ProProIleProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220
Qy      661  CTCTATAATTGAAAACTAACCATCCCCCTTTCTCGGAGGATCCCCAACGGCTCACGGGG 720
Db      221  LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240
Qy      721  TTGGTGGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGATGATGTCAACAGCTGCTG 780
Db      241  LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
Qy      781  CAGACACTCTTCACACCGGAGCGAGAGAGAATCTATTAGAGGCTAGAAAAAATGTT 840
Db      261  GlnThrLeuPheThrThrGluGluArgGluIleLeuLeuGluAlaArgLysAsnVal 280
Qy      841  CCTGGGCGGACGGGCGGACCCACCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA 900
Db      281  ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300
Qy      901  ACTGCCCGGTTGGGACTACAAACAGGCTGAAGTAGGAGAGCTTGAAAAATCTATGCG 960
Db      301  ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320
```

QY 961 CAGGCTCTGGTGGGGCTCTCGGGGGCGCTCAAGACGGCCCACTAATTTGGCTTAAGTA 1020  
Db GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgProThrAsnLeuAlaLysVal 340  
QY 1021 AGAGAAGTGATGACGGGACCGAATGAACCCCTCTGTTTCTTGAGAGGCTCTTGAA 1080  
Db ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
QY 1081 GCCTTCAGCGGTTACACCCCTTTTGATCCCTCAGAGGCGCCCAAAAGCCCTCAGTGCT 1140  
Db AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
QY 1141 TTGGCCTTTATAGGACAGCTAGCCTTGATATTAGAAAGAGCTTCAGAGACTGGAAGG 1200  
Db LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
QY 1201 TTACAGGAGGCTGAGTCTACGTGATCTAGTGAAGGAGGAGAGAAAGTATATTACAAAGG 1260  
Db LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrLysArg 420  
QY 1261 GAGACAGAGAAGAAAGGAAACAAAGAAAGAGAGAGAAAGAGAGAGAGAGAGAAAGA 1320  
Db GluThrGluGluGluArgAspGlnArgLysGluArgGluArgGluArgGluArg 440  
QY 1321 CGTAAATAAACGCAAGAGAAATTTGACTAAGATCTTGGCTGAGTGTTGAAGGAAA 1380  
Db ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys 460  
QY 1381 AGCAATACGGAAGAGAGAGATTTTAGGAATATTAGGTACAGCCCTAGACACTCAGG 1440  
Db SerAsnArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
QY 1441 AACCTGGCCATAGGACCCCACTCGCAAGACCAATGTGCATATTGTAAGAAAGAGGA 1500  
Db AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500  
QY 1501 CACTGGGCAAGAACTGCCCCAAGAGGAAACAAAGACCAAGATCTTAGCTCTAGAA 1560  
Db HisTrpAlaArgAspCysProLysLysGlyAsnLysGlyLeuLysValLeuAlaLeuGlu 520  
QY 1561 GAAGATAAAGAT 1572  
Db GlnAspLysAsp 524

## RESULT 8

Q90RM0\_9GAMR PRELIMINARY; PRT; 524 AA.  
ID Q90RM0\_9GAMR PRELIMINARY; PRT; 524 AA.  
AC Q90RM0;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Group specific antigen.  
GN Name=gag;  
OS Porcine endogenous type C retrovirus.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gammatetrovirus.  
OX NCBI\_TaxID=70540;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21256017; PubMed=11356953;  
RX DOI=10.1128/JVI.75.12.5465-5472.2001;  
RA Krach U., Fischer N., Czauderna F., Toenjes R.R.;  
RT "Comparison of replication-competent molecular clones of porcine  
endogenous retrovirus class A and class B derived from pig and human  
cells.";  
RL J. Virol. 75:5465-5472 (2001).  
DR EMBL; AJ293656; CAC39616.1; -; Genomic\_DNA.  
DR HSSP; P03332; 1A6B.  
DR SMR; Q90RM0; 7-98.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.

GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR003036; Gag\_p30.  
DR InterPro; IPR000840; G retro matrix.  
DR InterPro; IPR001878; ZnF\_CCHC.  
DR Pfam; PF01140; Gag\_MA; 1.  
DR Pfam; PF02093; Gag\_p30; 1.  
DR Pfam; PF00098; zf-CCHC; 1.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR SMART; SM00343; ZnF\_C2HC; 1.  
DR PROSITE; PS0158; ZnF\_CCHC; 1.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 524 AA; 59486 MW; 83B73BB3405C9D8A CRC64;

## Alignment Scores:

Pred. No.: 3,42e-164 Length: 524  
Score: 2745.00 Matches: 508  
Percent Similarity: 98.9% Conservative: 10  
Best Local Similarity: 96.9% Mismatches: 6  
Query Match: 96.4% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x Q90RM0\_9GAMR (1-524)

QY 1 ATGGGACAGACGGTGAAGACCCCTCTTAGTTTGAATCTCGACCATTTGACTTGAAGTTAAA 60  
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValLys 20  
QY 61 TCCAGGCTCATATTTGTCTAGTTTCAGTTTAAAGAGGACCTTCGGCAGACTTCTGTGTC 120  
Db 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysVal 40  
QY 121 TCTGAATGGCGCAGCATTCGATTTGGATGGCCATCAGAGGGGACCTTTAATTTCTGAGATT 180  
Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
QY 181 ATCTGCTGTTAAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCGATCAGAG 240  
Db 61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
QY 241 CCTTATATCTTACGTGCGCAAGATTTGGCAGAGGATCTCCGCCATGGTTAAACCATGG 300  
Db 81 ProTrpIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100  
QY 301 CTGAATAAGCAAGAACCCAGGTCCCGGAATCTTGCTCTCTTGAGAGAGAAAAACACAC 360  
Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
QY 361 TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTACCCGAGATTGAGGACCCAGGCT 420  
Db 121 SerAlaGluLysValLysProSerProHisIleTyrProGluIleGluGluProProAla 140  
QY 421 TGGCCGGAACCCCAATCTGTTCCCCACCCCTTATCTGGCAGAGGTCGCCGAGGGGA 480  
Db 141 TrpProGluGlnSerValProProProTrpProAlaGlnGlyAlaAlaArgGly 160  
QY 481 CCTTTTCCCTCTCTGGAGCTCCGGCGGTGGAGGACCTCTGCGAGGACTCGGAGCGGG 540  
Db 161 ProSerAlaProProGlyAlaProAlaValGluGlyProAlaAlaGlyThrArgSerArg 180  
QY 541 AGGGGCGCCACCCCGGAGCGGACAGAGATCGGACATTACCGCTGCGCAGCTACGGC 600  
Db 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyrGly 200  
QY 601 CCTCCACACACCGGGGGGCGAATTGACGCCCTCCAGTATTGGCCCTTTCTTCTGCAGAT 660  
Db 201 ProProIleProGlyGlnLeuGlnProLeuGlnTrpProPheSerSerAlaAsp 220  
QY 661 CTCTATAATTGGAAAACTAACCACTCCCTCTTCGAGGATCCCCAACCCCTCACCGGG 720  
Db 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
QY 721 TTGGTGGAGTCCCTTATGTTCTCTCACCGCTTCTTGGGATGATTCTCAACAGCTCTG 780

```
Db      241  LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
Qy      781  CAGACACTCTTCACACACCGAGGAGGAGAGAAATCTATTAGAGCGCTAGAAAAATGTT 840
Db      261  GlnThrLeuPheThrThrGluGluArgLysArgIleLeuLeuGluAlaAArgLysAsnVal 280
Qy      841  CCTGGGCGGAGCGGCGACCCACGCGGTTCGAAATGAGATTGACATCGGATTTCCCTTA 900
Db      281  ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300
Qy      901  ACTGCCCGCGGTGGGACTACACACACGCGCTGAAGGTAGGAGAGCTTGAAATCTATCGC 960
Db      301  ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320
Qy      961  CAGGCTCTGGTGGGCGGTCTCGGGGCGCCTCAGACGCGCCACTAAATTTGGCTAAGGTA 1020
Db      321  GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340
Qy      1021  AGAGAAGTGATGCAGGAGCCGAATGAACCCCTCTGTTTCTTTGAGAGGCTCTTGAA 1080
Db      341  ArgGluValMetGlnGlyProAsnGluProSerValPheLeuGluArgLeuMetGlu 360
Qy      1081  GCCTTCAGCGGTACACCCCTTTTGTATCCCACTCAGAGGCCCAAAAGCCCTCAGTGGCT 1140
Db      361  AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerMetAla 380
Qy      1141  TTGGCCTTTATAGACAGCTAGCCTTGATATATAGAAAGAGCTTCAGAGACTGGAAGGG 1200
Db      381  LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400
Qy      1201  TTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGCGCAGAGAAAGTATATTACAAAGG 1260
Db      401  LeuGlnGluAlaGluLeuHisAspLeuValLysGluAlaGluLysValTyrTyrLysArg 420
Qy      1261  CAGACAGAAGAAAGGAAACAAAGAAAGAGAGAGAGAGAGAAAGGAGGAGAAAGA 1320
Db      421  GluThrGluGluGluArgGluGlnArgLysGluArgGluArgGluArgGluArg 440
Qy      1321  CGTAATAACCGGCAAGAGAAGAAATTTGACTAAGATCTTGGCTCGAGCTGTTGAAGGGAAA 1380
Db      441  ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValAlaGluGlyLys 460
Qy      1381  AGCAATACGGAAGAGAGAGAGATTTTAGGAAATTTAGTCAGGCGCTCAGACAGTCAGGG 1440
Db      461  SerAsnArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480
Qy      1441  AACCTGGGCAATAGGACCCCACTCGACAAGGACCAATGTGCATATTTGTAAGAAGAGAGA 1500
Db      481  AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500
Qy      1501  CACTGGGCAAGGAAGTGGCCCAAGAGGGAACAAAGGACCAAGGATCTTAGCTCTAGAA 1560
Db      501  HisTrpAlaArgAspCysProLysLysGlyAsnLysGlyLeuLysValLeuAlaLeuGlu 520
Qy      1561  GAAGTAAAGAT 1572
Db      521  GluAspLysAsp 524

RESULT 9
ID      Q6UJZ5_9GAMR PRELIMINARY; PRT; 524 AA.
AC      Q6UJZ5;
DT      05-JUL-2004 (TrEMBLrel. 27, Created)
DT      05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE      Gag protein.
DE      Gag protein.
CN      Name=gag;
OS      Porcine endogenous retrovirus.
OC      Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC      Gamaretrovirus; 1-Mammalian type C virus group.
OX      NCBI_TaxID=61673;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
```

```
RX      PubMed=14963152; DOI=10.1128/JVI.78.5.2502-2509.2004;
RA      Scobie L., Taylor S., Wood J.C., Suling K.M., Quinn G., Meikle S.,
RA      Patience C., Schuurman H.J., Onions D.E.;
RT      "Absence of replication-competent human-tropic porcine endogenous
RT      retroviruses in the germ line DNA of inbred miniature Swine.";
RL      J. Virol. 78:2502-2509(2004).
DR      EMBL; AY368583; AAQ73933.1; -; Genomic_DNA.
DR      HSSP; P03322; 146B.
DR      SMR; Q6UJZ5; 7-98.
DR      GO; GO:0019028; C:viral capsid; IEA.
DR      GO; GO:0046872; F:metal ion binding; IEA.
DR      GO; GO:0003676; F:nucleic acid binding; IEA.
DR      GO; GO:0005198; F:structural molecule activity; IEA.
DR      GO; GO:0019088; P:viral assembly; IEA.
DR      InterPro; IPR003036; Gag_p30.
DR      InterPro; IPR000840; G_retro_matrix.
DR      InterPro; IPR001878; Znf_CCHC.
DR      Pfam; PF01140; Gag_MA; 1.
DR      Pfam; PF02093; Gag_p30; 1.
DR      Pfam; PF00098; zf-CCHC; 1.
DR      PRINTS; PR00939; C2HCZNFINGER.
DR      SMART; SM00343; Znf_C2HC; 1.
DR      PROSITE; PS0158; ZF_CCHC; 1.
DR      Metal-binding; Zinc; Zinc-finger.
SQ      SEQUENCE 524 AA; 59525 MW; B2AF9A672406B8EC CRC64;

Alignment Scores:
Pred. No.:      6,19e-163      Length:      524
Score:          2725.00      Matches:      507
Percent Similarity: 98.5%      Conservative: 9
Best Local Similarity: 96.8%      Mismatches: 8
Query Match:      95.7%      Indels: 0
DB:              2              Gaps: 0

US-10-723-552-3_COPY_585_2156 (1-1572) x Q6UJZ5_9GAMR (1-524)
Qy      1  ATGGGACAGAGCGGTGACGACCCCTCTAGTTTGACTCTCGACCATTGAGCTGAAGTTAAA 60
Db      1  MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValLys 20
Qy      61  TCCAGGGCTCATAAATTTGTCAAGTTTCAGGTTAAGAGGAGACCTTGGCAGACTTTCTGTGTC 120
Db      21  SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysVal 40
Qy      121  TCTGAATAGCGCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAAATCTCGAGATT 180
Db      41  SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60
Qy      181  ATCTCTGGCTGTTAAGCAGATTATTTTTCAGACTCGACCCGGCTCTCATCCCGATCAGGAG 240
Db      61  IleuAlaValLysAlaIleIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80
Qy      241  CCCTATATCTCTACGTGCGCAAGATTGTCAGAGGATCTCCCGCATGTTGGTTTAAACCATGG 300
Db      81  ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
Qy      301  CTGAATAGCCCAAGAACCCAGGTCCTCCCGAATCTCGCTCTTGGCTTGGAGAGAAAACAAACAC 360
Db      101  LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120
Qy      361  TCGGCTGAAGAAGTCAGGCCCTCTCCATATCTACCCCGAGATTGAGGAGCCACCGGCT 420
Db      121  SerAlaGluLysValLysProSerProHisIleTyrProGluIleGluGluProProAla 140
Qy      421  TGGCGGGAACCCCAATCTGTTTCCCGCACCCCTTATCTGGCACAGGGTTCGCGGAGGGA 480
Db      141  CysProGluProGlnSerValProProProTyrProAlaGlnGlyAlaAlaArgGly 160
Qy      481  CCCTTTGCCCTCTCTGGAGCTCCGCGGTGAGGAGACTGTCTGAGGAGACTCGGAGCCGG 540
Db      161  ProSerAlaProProGlyAlaProAlaValGluGluProAlaAlaGlyThrArgSerArg 180
Qy      541  AGGGCGGCCACCCCGGAGCGGACAGAGATCGGACATTACCGCTCGCACGTACGGC 600
```

Db 181 ArgGlyAlaThrProGluArgThrAspGluLeuAlaThrLeuProLeuArgThrTyrGly 200  
Qy 601 CTCTCCACACCGGGGGGCAATTTGAGCCCTCCAGTATTGGCCCTTTTCTTCTGACAGT 660  
Db 201 ProProlleProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
Qy 661 CTCTATAATTGGAATAACCATCCCTTTCTCGAGGATCCCAACGCTCACCGGG 720  
Db 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
Qy 721 TTGTGGAGTCCCTATTGCTCTCCACAGCTACTTGGGATGATTGTCAACAGTCTG 780  
Db 241 LeuMetGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeu 260  
Qy 781 CAGACACTCTTCAACACGAGGAGCGAGAGAATTTCTATTAGAGGCTAGAAAAATGTT 840  
Db 261 GlnThrLeuPheThrGluGluArgGluArgGluLeuLeuAlaArgLysAsnVal 280  
Qy 841 CTGGGGCGGACGGGCGACCCACGCGGTGCAAAATGAGATTGACATGGGATTTCCCTTA 900  
Db 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluLeuAspMetGlyPheProLeu 300  
Qy 901 ACTGCCCCGCTGGGACTCAACACGCTGAAGTAGGGAGAGCTTGAATCTATCCG 960  
Db 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
Qy 961 CAGGCTCTGTGGGGGTCTCCGGGGCCCTCAAGACGGCCCACTAATTTGGCTTAAGTA 1020  
Db 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
Qy 1021 AGAAGTGTAGCGGGACCGAATGAACCCCTCTGTTTCTTGAGAGGCTCTTGAA 1080  
Db 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
Qy 1081 GCCTTCAGGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAGCCCTCAGTGGT 1140  
Db 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
Qy 1141 TTGGCTTTATAGACAGTACGCTTGGATATTAGAAGAGCTTCAGAGACTCGAAGG 1200  
Db 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
Qy 1201 TTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAGGAGAGAAAGTATATTACAAAGG 1260  
Db 401 LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg 420  
Qy 1261 GAGACAGAGNAGAAAGGAACAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 1320  
Db 421 GluThrGluGluGluArgGluGlnArgLysGluArgGluArgGluArgGluArg 440  
Qy 1321 CGTAATAAAGCGCAAGAGAGAAATTTGACTAAGATCTTGGCTGCAGTGGTTGAAGGAAA 1380  
Db 441 ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaValGluGlyLys 460  
Qy 1381 AGCAATACGGAAGAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTTAGACAGTCAAGG 1440  
Db 461 SerAsnArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
Qy 1441 AACTGGGCAATAGGACCCACTCGACAGGACCAATGTGCATATTGTAAGAAGAGGA 1500  
Db 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500  
Qy 1501 CACTGGGCAAGGAACTGCCCAAGAGGAAACAAAGGACCAAGGACCTTAGCTCTAGAA 1560  
Db 501 HisTrpAlaArgAspCysProLysLysGlyAsnLysGlyLeuLysGluIleAlaLeuGlu 520  
Qy 1561 GAAGATAAAGAT 1572  
Db 521 GluAspLysAsp 524

RESULT 10  
Q6UJZ2\_9GAMR

ID Q6UJZ2\_9GAMR PRELIMINARY; PRT; 524 AA.  
AC Q6UJZ2;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE gag protein.  
GN Name=gag;  
OS Porcine endogenous retrovirus.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gammatetravirus; 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=14963152; DOI=10.1128/JVI.78.5.2502-2509.2004;  
RA Scobie L., Taylor S., Wood J.C., Suling K.M., Quinn G., Meikle S.,  
RA Patience C., Schuurman H.J., Onions D.E.;  
RT "Absence of replication-competent human-tropic porcine endogenous  
RT retroviruses in the germ line DNA of inbred miniature swine.";  
RL J. Virol. 78:2502-2509(2004).  
DR EMBL; AY368586; AAQ73935.1; -; Genomic\_DNA.  
DR HSSP; P03332; 1A6B.  
DR SMR; O6UJZ2; 7-98.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR003036; Gag\_P30.  
DR InterPro; IPR000840; G\_Retro\_matrix.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF01140; Gag\_MA; 1.  
DR Pfam; PF02093; Gag\_P30; 1.  
DR Pfam; PF00098; zf\_CCHC; 1.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR SMART; SM00343; Znf\_C2HC; 1.  
DR PROSITE; PSS0158; zf\_CCHC; 1.  
KW Metal-binding; Zinc; Zinc-finger.  
SQ SEQUENCE 524 AA; 59349 MW; 1F7CD5A5B8F13F8 CRC64;

Alignment Scores:  
Pred. No.: 7,15e-163 Length: 524  
Score: 2724.00 Matches: 507  
Percent Similarity: 98.1% Conservative: 7  
Best Local Similarity: 96.8% Mismatches: 10  
Query Match: 95.6% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3\_copy\_585\_2156 (1-1572) x Q6UJZ2\_9GAMR (1-524)

Qy 1 ATGGGACAGACGGTGCAGCCCTCTTGTAGTTTGAATCTCGACCATTCGACTGAAAGTTAAA 60  
Db 1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpThrGluValys 20  
Qy 61 TCAGGGCTCATTAATTTGTCTAGTTTCAGTTAAGAGGACCTTGGCAGACTTCTGTGTC 120  
Db 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysVal 40  
Qy 121 TCTGAATGGCCGACATTCGATGTGGATGGCCATCAGAGGGACCTTTAATCTGAGATT 180  
Db 41 SerGlnTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
Qy 181 ATCCTGGCTTTAAAGCAGTTATTTTTCAGACTGGACCCCGCTCTCATCCGATCAGAG 240  
Db 61 IleLeuAlaValLysAlaIleIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
Qy 241 CCCTATATCTTACGTGGCAAGATTGGCAGAGATCTCCGCCATCGGTTAAACCATGG 300  
Db 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100  
Qy 301 CTGAATTAAGCAAGAACCCAGGTCGCCGAATTTCTGGCTCTTGGAGAGAAAACACAC 360  
Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120



```
Qy 361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGAGCCACCGGCT 420
Db 121 SerAlaGluValLeuValProSerProHisIleLeuProGluIleGluGluProProAla 140
Qy 421 TGGCGGGAACCCCAATCTGTTCCCCACCCCTTATCTGGCACAGGGTGCAGGGGGA 480
Db 141 GlyProGluProGlnSerValProProProProTyProAlaGlnGlnAlaAlaArgGly 160
Qy 481 CCTTTGGCCCTCTCGAGCTCGGCGGTGAGGACCTGCTGCAGGAGCTCGAGCGGG 540
Db 161 ProSerAlaProProGlyAlaProAlaValGluGluProAlaAlaGlyThrArgSerArg 180
Qy 541 AGGGCGCCACCCGAGCGGACAGACGAGATCGACATTACGCTCGCGCACGTACGGC 600
Db 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTrpGly 200
Qy 601 CCTCCACACCGGGGGGCAATTGTCAGCCCTCCAGTATTCGCCCTTTCTTCGCGAT 660
Db 201 ProProIleProGlyGlyGlnLeuGlnProLeuGlnTyTrpProPheSerSerAlaAsp 220
Qy 661 CTCTATATTTGGAACCTAACATCCCTCTTCGAGAGATCCCCAAAGCTCAGCGGG 720
Db 221 LeuTyAsnTrpLeuThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240
Qy 721 TTGGTGGAGTCCCTTATGTTCTCTCACCAGCTACTTCGGATGATTGTCAACAGCTGCTG 780
Db 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
Qy 781 CAGACACTCTTTCACACCGGAGCGAGAGAGAAATCTTATTAGAGGCTAGAAAAATGTT 840
Db 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaAlaArgLysAsnVal 280
Qy 841 CTTGGGCGGCGAGCGGCGACCCACGCGGTGTCAAATGATGATGATGGGATTTCCCTTA 900
Db 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300
Qy 901 ACTGCCCGGTGGGACTACAAACGCTCAAGGTAGGAGAGCTTCAAAATCTATCGC 960
Db 301 ThrArgProGlyTrpAspTyAsnThrAlaGluGlyArgGlySerLeuLysIleTyArg 320
Qy 961 CAGGCTCTGGTGGCGGGTCTCGGGGCGCCTCAAGACGCGCCACTAAATTTGGCTAAGGTA 1020
Db 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340
Qy 1021 AGAGAAGTATGTCAGGAGCGGAATGAACCCCGCTCTGTTTCTTGACAGGCTCTTGAA 1080
Db 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360
Qy 1081 GCCTTCAGCGGTACACCCCTTTGATCCACCTCAGAGGCCCAAAAGCCCTCAGTGGCT 1140
Db 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380
Qy 1141 TTGGCCTTTATAGGACAGTCAGCCTTGGATATTAGAAAAGCTTTCAGAGCTCGAAGGG 1200
Db 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400
Qy 1201 TTACAGAGGCTGAGTTACGTACTAGTAAAGGAGGACAGAAAGTATATTACAAAGG 1260
Db 401 LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyTrpLysArg 420
Qy 1261 GAGACAGAGAAGAAAGGGAAACAAAGAAAAGAGAGAGAGAAAGGAGGAGGAGAAAGA 1320
Db 421 GluThrGluGluGluArgAspGlnArgLysGluArgGluArgGluGluGluArg 440
Qy 1321 CGTATAAACCGCAAGAGAGAATAATTGACTTTGGCTGCGAGTGGTGGTGAAGGGAAA 1380
Db 441 ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValAlaGluGlyLys 460
Qy 1381 AGCAATACGGAAGAGAGAGAGATTTTAGGAAAATTTAGTCTAGGCCCTTAGACAGTCAGG 1440
Db 461 SerAsnArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480
Qy 1441 AACCTGGGCAATAGGACCCCTCTCGACCAAGGACCAATGTGCATATTGTAAAGAAAGAGGA 1500
```

```
Db 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrcysLysGluLysGly 500
Qy 1501 CACTGGCGAGGAAGTCTCCCAAGAGAGGAAACAAAGGACCAAGATCTCTAGCTCTAGAA 1560
Db 501 HistTrpAlaArgAspCysProLysGlyAsnLeuGlyThrGluAsnLeuThrLeuGlu 520
Qy 1561 GAAGATAAAGAT 1572
Db 521 GluAspLysAsp 524

RESULT 11
Q9NLT7_PIG PRELIMINARY; PRT; 664 AA.
AC Q9NLT7_PIG PRELIMINARY; PRT; 664 AA.
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DB Gag-pol (Fragment).
OS Sus scrofa domestica (domestic pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OX NCBI_TaxID=9825;
RN [1]
NUCLEOTIDE SEQUENCE.
RC STRAIN=Breed Duroc x German Landbreed;
RA Blusch J.H., Seelmeier S., von der Helm K.;
RT "Evidence for extreme sequence homogeneity of the PERV A/B 5'
RT LTR/leader/gag/prot region by predicted targeted amplification from a
RT Du XL pig.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF147808; AAE65926.1; -; Genomic_DNA.
DR HSSP; Q9WUP4; 1BW4.
DR SMR; Q9NLT7; 7-98.
DR MEROPS; A02.020; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0019068; P:viral assembly; IEA.
DR InterPro; IPR003036; Gag_p30.
DR InterPro; IPR000840; G_retro_matrix.
DR InterPro; IPR001995; Peptidase_A2_cat.
DR InterPro; IPR001969; Pept_Asp_AS.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01140; Gag_MA; 1.
DR Pfam; PF02093; Gag_p30; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
DR PROSITE; PS0158; zf_CCHC; 1.
FT CHAIN 1 524 gag protein.
FT CHAIN 525 >664 aspartic protease.
FT NON_TER 664
SQ SEQUENCE 664 AA; 75259 MW; 188FA83D39FE2F02 CRC64;

Alignment Scores:
Pred. No.: 1,15e-161 Length: 664
Score: 2705.00 Matches: 497
Percent Similarity: 97.9% Conservative: 16
Best Local Similarity: 94.8% Mismatches: 11
Query Match: 95.0% Indels: 0
DB: 2 Gaps: 0

US-10-723-552-3_copy_585_2156 (1-1572) x Q9NLT7_PIG (1-664)
Qy 1 ATGGACAGACGCGTGACGACCCCTCTTAGTTTGACTCTCGACCATTCGAGCTGAAGTTAAA 60
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20
```

QY 61 TCCAGGGCTCATATTTGTCAGTTTCAGGTTTAAAGAGGACCTTGGCAGACTTTCGTGTC 120  
Db |||||  
21 SerArgAlaHisAsnLeuSerValGlnValIysGlyProTrpGlnThrPheCysAla 40  
QY 121 TCTGAATGGCGGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAAATCTGAGATT 180  
Db |||||  
41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
QY 181 ATCTGGCTGTAAAGCAGTATTATTTTTCAGACTCGACCCGGCTCTCATCCCGATCAGGAG 240  
Db |||||  
61 IleLeuAlaValIysAlaIleIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
QY 241 CCCTATATCTTACGTGGCAAGATTGGCAGAGGATCTCCGCCATGGGTTTAAACCATGG 300  
Db |||||  
81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValIysProTrp 100  
QY 301 CTGAATGAAGCAAGAACCCAGGTCCCGGAATTCGGCTCTTGGAGAGAGAAACAAACAC 360  
Db |||||  
101 LeuAsnIysProArgIysProGlyProArgIleLeuAlaLeuGlyGluIysAsnLysHis 120  
QY 361 TCGGCTGAAAAGTCAAGCCCTCTCCTCATATCTACCCGAGATTGAGGAGCCACCGCT 420  
Db |||||  
121 SerAlaGluIysValGluProSerProArgIleTyrProGluIleGluGluProThr 140  
QY 421 TGGCCGGAACCCCAATCTGTTCCCCACCCCTTATCTGGCACAGGTTGCCGAGGGGA 480  
Db |||||  
141 TrpProGluProGlnProValProProProProTyrProAlaGlnGlyAlaValArgGly 160  
QY 481 CCCTTTCCTCTCTGAGCTCCGGCGGTGGAGGACCTGCTGAGGAGCTCGGAGCGG 540  
Db |||||  
161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180  
QY 541 AGGGGCCCCACCCCGAGCGACAGACGAGATCCGACATTCACCGCTGCGCAGCTACGGC 600  
Db |||||  
181 ArgGlyAlaThrProGluIysGlnThrAspGluIleAlaLeuProLeuArgThrTyrGly 200  
QY 601 CCTCCACACCGGGGGCCCAATTGTCAGCCCTCCAGTATTGGCCCTTTCTTCTGAGAT 660  
Db |||||  
201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
QY 661 CTCTATAATTTGGAATACTACCATCCCGCTTCTCGGAGGATCCCAACCGCTCACGGG 720  
Db |||||  
221 LeuTyrAsnTrpIysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
QY 721 TTGCTGGAGTCCCTTATGTTCTCTCACAGCTACTTGGGATGATTCCTCAACAGCTCTG 780  
Db |||||  
241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeu 260  
QY 781 CAGACACTCTTCAACCGAGGCGAGAGAGAAATCTATTAGAGGCTAGAAAATGTT 840  
Db |||||  
261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280  
QY 841 CCTGGGCGCCAGCGGCGACCCAGCGGTTGCAAAATGAGATTGATGAGGATTCCTTTA 900  
Db |||||  
281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
QY 901 ACTCGCCCGGTTGGGACTACACACCGCTGAGGTAGGAGGACTTGAATCTATCCG 960  
Db |||||  
301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuIysIleTyrArg 320  
QY 961 CAGGCTCTGCTGGCGGTCTCCGGGGCGCTCAAGACGGCCCTAATTTGGCTTAAGGTA 1020  
Db |||||  
321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgProThrAsnLeuAlaIysVal 340  
QY 1021 AGAAGTGTAGCAGGACCGAATGAACCCCTCTCTTTTCTTGAGAGGCTCTTGAA 1080  
Db |||||  
341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
QY 1081 GCCTTCAGCGGTACACCCCTTTTGATCCCACTCAGAGGCCCAAAAGCTCAGTGGCT 1140  
Db |||||  
361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380

QY 1141 TTGGCCCTTTATAGACAGTCTAGCCTTGATATTAGAAGAAGCTTCAGAGACTGGAAGG 1200  
Db |||||  
381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgIysLysLeuGlnArgLeuGluGly 400  
QY 1201 TTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAGGAGAGAAAGTATATATACAAAGG 1260  
Db |||||  
401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluIysValTyrTyrArg 420  
QY 1261 GAGACAGAAGAAGGAGCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db |||||  
421 GluThrGluGluGluIysGluGlnArgLysGluIysGluArgGluGluArgGluGlu 440  
QY 1321 CGTAATAAAGCGCAAGAGAAATTTGACTAAGATCTTGGCTGAGTGGTTTGAAGGAAA 1380  
Db |||||  
441 ArgAspArgArgGlnGluLysAsnLeuThrIysIleLeuAlaValValGluGlyLys 460  
QY 1381 AGCAATACGGAAGAGAGAGAGATTTTAGGAAATTAGTTCAGGCCCTAGACAGTCAGG 1440  
Db |||||  
461 SerSerArgGluArgGluArgPheArgLysIleArgSerGlyProArgGlnSerGly 480  
QY 1441 AACCTGGCAATAGGACCCCTCGACAAGGACCAATGTGCATATTGTAAGAAGAGAGA 1500  
Db |||||  
481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluIysGly 500  
QY 1501 CACTGGCAAGGAACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
Db |||||  
501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520  
QY 1561 GAAGATTAAGAT 1572  
Db |||||  
521 GluAspLysAsp 524

RESULT 12  
Q8J4V8\_9GAMR  
ID Q8J4V8\_9GAMR PRELIMINARY; PRT; 1718 AA.  
AC Q8J4V8;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Porcine endogenous retrovirus A.  
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=194958;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22172947; PubMed=12185278;  
RA Bartosch B., Weiss R.A., Takeuchi Y.;  
RT "PCR-based cloning and immunocytological titration of infectious  
RL J. Gen. Virol. 83:2231-2240(2002).  
DR EMBL; AY099323; AAM29192.1; -; Genomic\_DNA.  
DR HSP; P03355; I160.  
DR SMR; Q8J4V8; 7-98.  
DR MEROPS; A02.020; -.  
DR GO; GO:0019028; C:Viral capsid; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR003036; Gag\_p30.  
DR InterPro; IPR000840; G retro matrix.  
DR InterPro; IPR001995; Peptidase\_A2\_cat.  
DR InterPro; IPR001989; Pept\_Asp\_AS.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.

DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF01140; Gag\_MA; 1\_CCHC.  
 DR Pfam; PF02093; Gag\_P30; 1.  
 DR Pfam; PF00075; RnaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT; 1; 1.  
 DR Pfam; PF00098; zf\_CCHC; 1.  
 DR SMART; SM00343; Znf\_C2HC; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS50879; RNASE H; 1.  
 DR PROSITE; PS50158; zf\_CCHC; 1.  
 KW Polyprotein.  
 SQ SEQUENCE 1718 AA; 193115 MW; 4E86215286C59246 CRC64;

Alignment Scores:  
 Pred. No.: 1,26e-161 Length: 1718  
 Score: 2705.00 Matches: 497  
 Percent Similarity: 97.9% Conservative: 16  
 Best Local Similarity: 94.8% Mismatches: 11  
 Query Match: 95.0% Indels: 0  
 DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x Q8J4V8\_9GAMR (1-1718)

QY 1 ATGGAGCAGCGGTGACGACCCCTCTAGTTGACTCTCGACCATAGGCTGAAGTTAAA 60  
 DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20

QY 61 TCCAGGGCTCATATTTCTCAGTTTCAGGTTAAGAGGACCTTGGCAGCTTTCTGTGTC 120  
 DB 21 SerArgAlaHisAsnLeuSerValGlnValLysGlyProTrpGlnThrPheCysAla 40

QY 121 TCTGAATGCCGACATTCGATGTTGGATGGCCATCAGAGGGACCTTTAAATTCGAGATT 180  
 DB 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluile 60

QY 181 ATCTGGCTGTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCGATCAGGAG 240  
 DB 61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80

QY 241 CCCTATATCTTACTGTGCAAGATTGGCAGAGATCTCCGCCATGGTGTAAACCATGG 300  
 DB 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpTrpValLysProTrp 100

QY 301 CTGAATAAGCAAGAACCCAGGTCCTCCGAATTTCTGGCTCTTGGAGAGAAAACAAACAC 360  
 DB 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120

QY 361 TCGGCTGAAAAGTCAGCCCTCTCCTCATATCTACCCCGAGATTGAGGACCCAGGCT 420  
 DB 121 SerAlaGluLysValGluProSerProArgIleTyrProGluileGluGluProProThr 140

QY 421 TGGCCGGAACCCCAATCTGTTCCCCACCCCTTATCTGGCAGACGGTGGCCGAGGGGA 480  
 DB 141 TrpProGluProGlnProValProProProTrpProLysProAlaGlnGlyAlaValArgGly 160

QY 481 CCCTTTGCCCCCTCTCTGGAGCTCCGCGGTGGAGGACCTGTCTGAGGAGCTCGAGCCGG 540  
 DB 161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180

QY 541 AGGGGCGCCACCCCGGAGCGGACAGACGAGATCGGCATATTACGCTCGGCACGTACGGC 600  
 DB 181 ArgGlyAlaThrProGluArgThrAspGluileAlaIleLeuProLeuArgThrTyrGly 200

QY 601 CCTCCACACCGGGGGGCAATTGAGCCCTTCAGATTGGCCCTTTCTTCGACAT 660  
 DB 201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220

QY 661 CTCTATAATGGAAAACCAATCCATCCCTTTCTCGGAGGATGCCCAACGCTCAGCGGG 720  
 DB 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240

QY 721 TTGCTGGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGATGATTGTCAACAGCTGCTG 780  
 DB 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspPheCysGlnGlnLeuLeu 260

QY 781 CAGACACTCTTTCACAAACCGAGGACGAGAGAGAAATTTCTATTAGAGGCTAGAAAAATGTT 840  
 DB 261 GlnThrLeuPheThrThrGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280

QY 841 CCTGGGCGGACGGGCGACCCACCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA 900  
 DB 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300

QY 901 ACTGCGCCCGTTCGGGACTACAACACGCTCAAGCTAGGAGAGCTTCAAAATCTATCGC 960  
 DB 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320

QY 961 CAGCTCTGTGTGGCGGTCTCCGGGCGCTCTCAAGCGGCCCACTAAATTTGGCTAAGTA 1020  
 DB 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340

QY 1021 AGAAGATGATGCAGGACCGAATGAACCCCTCTGTTTCTTGAGAGCTCTTGAA 1080  
 DB 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360

QY 1081 GCCTTTCAGGGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAGCCCTCAGTGGCT 1140  
 DB 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380

QY 1141 TTGCGCTTTATAGACACAGTCTCAGCCTTGGATATTAGAAGAAAGCTTTCAGAGACTGGAAGG 1200  
 DB 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400

QY 1201 TTACAGAGGCTGAGTTACGTGATCTAGTGAAGAGGACGAGAAAGTATATTACAAAGG 1260  
 DB 401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg 420

QY 1261 GAGACAGAGAAGAAAGGGAACAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 DB 421 GluThrGluGluGluLysGluGlnArgLysGluLysGluArgGluArgGluGluArg 440

QY 1321 CGTAATAAACCGCAAGAGAGAAATTTGACTAAGATCTTGGCTGCAGTGGTTGAAGGGAAA 1380  
 DB 441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys 460

QY 1381 AGCAATACGGAAGAGAGAGAGATTTTAGAAAATTTAGTTCAGGCCCTAGACAGTCAGG 1440  
 DB 461 SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480

QY 1441 AACCTGGGCATAGGACCCCACTCGACAAGACCAATGTGCATATTGTAAAGAAAGAGA 1500  
 DB 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500

QY 1501 CACTGGGCAAGAACTGCCCCAAGAGGAAACAAAGGACCAAGGATCTTAGTCTTAGAA 1560  
 DB 501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520

QY 1561 GAAGATTAAGAT 1572  
 DB 521 GluAspLysAsp 524

RESULT 13  
 Q8J4V6\_9GAMR  
 ID Q8J4V6\_9GAMR PRELIMINARY; PRT; 1720 AA.  
 AC Q8J4V6;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Porcine endogenous retrovirus B.  
 OC Viruses; Retrovirdae; Gammaretrovirus;  
 OC 1-Mammalian type C virus group.  
 OX NCBI\_TaxID=194959;

[1]  
 RN NUCLEOTIDE SEQUENCE.  
 RP MEDLINE=22172947; PubMed=12185278;  
 RA Bartosch B., Weiss R.A., Takeuchi Y.;  
 RT "PCR-based cloning and immunocytological titration of infectious  
 RL J. Gen. Virol. 83:2231-2240(2002).  
 DR EMBL: AY093324; AAC29194.1; -; Genomic\_DNA.  
 DR HSSP: P03355; 116J.  
 DR SMR: Q8J4V6; 7-98.  
 DR MEROPS: A02.020; -;  
 DR GO: GO:0019028; C:viral capsid; IEA.  
 DR GO: GO:0004190; P:aspartic-type endopeptidase activity; IEA.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR GO: GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO: GO:0003723; F:RNA binding; IEA.  
 DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO: GO:0005198; F:structural molecule activity; IEA.  
 DR GO: GO:0006310; P:DNA recombination; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO: GO:0019068; P:viral assembly; IEA.  
 DR InterPro: IPR003036; Gag\_p30.  
 DR InterPro: IPR000840; G\_retro.matrix.  
 DR InterPro: IPR001995; Peptidase\_A2\_cat.  
 DR InterPro: IPR001969; Pept\_Asp\_AS.  
 DR InterPro: IPR002156; RNaseH.  
 DR InterPro: IPR001584; Rve.  
 DR InterPro: IPR000477; RVTse.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF01140; Gag\_MA; 1.  
 DR Pfam: PF02093; Gag\_p30; 1.  
 DR Pfam: PF00075; RNaseH; 1.  
 DR Pfam: PF00665; rve; 1.  
 DR Pfam: PF00077; RVP; 1.  
 DR Pfam: PF00078; RVT; 1.  
 DR Pfam: PF00098; zf\_CCHC; 1.  
 DR SMART: SM00343; Znf\_C2HC; 1.  
 DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE: PS01175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE: PS00879; RNASE\_H; 1.  
 DR PROSITE: PS0158; ZF\_CCHC; 1.  
 KW POLYPEPTIDE.  
 SQ SEQUENCE 1720 AA; 193393 MW; 7F9DCC2F5B834D8A CRC64;  
 Alignment Scores:  
 Pred. No.: 1.26e-161 Length: 1720  
 Score: 2705.00 Matches: 497  
 Percent Similarity: 97.9% Conservative: 16  
 Best Local Similarity: 94.8% Mismatches: 11  
 Query Match: 95.0% Indels: 0  
 DB: 2 Gaps: 0  
 US-10-723-552-3\_COPY\_585\_2156 (1-1572) x Q8J4V6\_9GMR (1-1720)  
 QY 1 ATGGGACAGCGGTGACGACCCCTCTTAGTTGACTCTCGACCAATGGACTGAAGTTAAA 60  
 DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20  
 QY 61 TCCAGGCTCATATTTGTCAGTTTCAAGTAAGGACCTTGGCAGACTTCTGTGTC 120  
 DB 21 SerArgAlaHisAsnLeuSerValGlnValValysGlyProTrpGlnThrPheCysAla 40  
 QY 121 TCTGAATGGCGGACATTCGATGTTGGATGGCATTCAGAGGGGACCTTTAATCTCGAGATT 180  
 DB 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
 QY 181 ATCTGGCTGTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCCGACTCAGGAG 240  
 DB 61 IleLeuAlaValysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
 QY 241 CCTTATATCTTACGTGGCAGATTTCGCGAGAGATCTCCGCGCATGGTTAAACCATGG 300

DB 81 ProTyrlleLeuThrTrpGlnAspLeuAlaGluAspProProProProTrpValysProTrp 100  
 QY 301 CTGAATAAGCCCAAGAACCCAGGTCCCGCAATTCCTGGCTCTTGGAGAGAAAAACACAC 360  
 DB 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGlyLeuAsnLysHis 120  
 QY 361 TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGAGCCACCGCT 420  
 DB 121 SerAlaGluLysValGluProSerProArgIleTyProGluIleGluGluProProThr 140  
 QY 421 TGGCCGGAACCCCAATCTGTTCCTCCCAACCCCTTATCTGGCACAGGTGCCCGAGGGA 480  
 DB 141 TrpProGluProGlnProValProProProProTyProAlaGlnGlyAlaValArgGly 160  
 QY 481 CCTTTTGGCCCTCTCTGGAGCTCCCGCGTGGAGGACCTGTCTGAGGAGCTCGGAGCCGG 540  
 DB 161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180  
 QY 541 AGGGCGCCACCCCGAGCGGACAGACGAGATCCGACATTACCGCTGCGCACCTACCGC 600  
 DB 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyGly 200  
 QY 601 CCTCCACACCGGGGGCCCAATTGACGCCCCCTCCAGTATTCGCCCTTTCTCTGTCAGAT 660  
 DB 201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyTrpProPheSerSerAlaAsp 220  
 QY 661 CTCTATAATTGAAAAAATAACCATCCCTTTCTCGAGGATCCCAACGCTCACGGGG 720  
 DB 221 LeuTyraTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
 QY 721 TTGGTGGAGTCCCTTATGTTCTCTCACGACCTACTTGGGATGATTGTCAACAGCTGCTG 780  
 DB 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnLeuLeu 260  
 QY 781 CAGACACTCTTCAACCGGAGCGGACGAGAGAAATCTATTAGAGCTAGAAAAATGTT 840  
 DB 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280  
 QY 841 CCTGGGCGCCAGCGGCGCACCGCGTTCGCAAAATGAGATTGATGATGGGATTTCCCTTA 900  
 DB 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
 QY 901 ACTCGCCCCGTTGGGACTACAACACCGCTGAAGGTAGGAGAGCTTGAATAATCTATCGC 960  
 DB 301 ThrArgProGlyTrpAspTyraThrAlaGluGlyArgGluSerLeuLysIleTyArg 320  
 QY 961 CAGGCTCTGTGGCGGTCTCCGGGCGCTCAAGAGCGCCCACTAATTTGGCTTAAGTA 1020  
 DB 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
 QY 1021 AGAAGTGTATGAGGAGCGACCGAATGAACCCCTCTCTTTCTTGTAGAGGCTCTTGAA 1080  
 DB 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
 QY 1081 GCCTTCAGGCGGTACACCCCTTTTGTATCCCACTCAGAGCGCCCAAAAGCCCTCAGTGCT 1140  
 DB 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
 QY 1141 TTGGCTTTTATAGACAGTCTGAGCTTGGATATTAGAAAGAGCTTCAGAGACTCGAAGGG 1200  
 DB 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
 QY 1201 TTACAGGAGCTCAGTTTACGTGATCTAGTGAAGGAGGACAGAGAAAGTATATTACAAAGG 1260  
 DB 401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyTrpArgArg 420  
 QY 1261 GAGACAGAAGAAGGAGGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 DB 421 GluThrGluGluGluLysGluGlnArgLysGluLysGluArgGluGluArgGluArg 440  
 QY 1321 CGTAATAACCGGACAGAGAAATTTGACTTAAGATCTTGGCTGTCAGTGGTTGAAGGAAA 1380  
 DB 441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys 460

```
QY 1381 AGCAATCGGAAGAGAGAGATTTTAGGAAATTTAGGTGAGCCCTAGACAGTCAGGG 1440
Db |||:::|||||
QY 461 SerSerArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGluArgGlu 480
Db |||:::|||||
QY 1441 AACCTGGGCAATAGAGACCCCTGACAAAGGACCAATGTGCATATTTGTAAGAAGAGGA 1500
Db |||:::|||||
QY 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500
Db |||:::|||||
QY 1501 CACTGGGCAAGAACTGCCCCCAAGGGAACAAAGGACCAAGGATCCTAGCTCTAGAA 1560
Db |||:::|||||
QY 501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520
Db |||:::|||||
QY 1561 GAAGATAAGAT 1572
Db |||:::|||||
QY 521 GluAspLysAsp 524
Db |||:::|||||

RESULT 14
Q9Q1X4_9GAMR
ID Q9Q1X4_9GAMR PRELIMINARY; PRT; 2378 AA.
AC Q9Q1X4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Type C proviral gag, pol and env genesand LTR (class A, clone
DE 421.
OS Porcine endogenous retrovirus.
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC Gamaretrovirus; 1-Mammalian type C virus group.
OX NCBI_TaxID=61673;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-Type C;
RA Czauderna F., Fischer N., Boller K., Krach U., Kurth R., Toenjes R.R.;
RT "Molecular Characterization of Human-tropic and Replication-competent
RT Porcine Endogenous Retroviruses."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ133817; CAB65340.1; -; GenomicDNA.
DR HSRF; P03385; IMOF.
DR SMR; Q9Q1X4; 7-98, 2227-2279.
DR MEROPS; A02.020; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0019068; P:viral assembly; IEA.
DR InterPro; IPR008981; FmULV-recept-bind.
DR InterPro; IPR003036; Gag_p30.
DR InterPro; IPR000840; G_retro-matrix.
DR InterPro; IPR001995; Peptidase_A2_cat.
DR InterPro; IPR000223; Peptidase_S26A.
DR InterPro; IPR001969; Pept_Asp_AS.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVtse.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF01140; Gag_MA; 1.
DR Pfam; PF02093; Gag_p30; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT; 1.
DR Pfam; PF00429; TLV_coat; 1.
DR Pfam; PF00098; zf_CCHC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
```

```
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
DR PROSITE; PS0879; RNASE_H; 1.
DR PROSITE; PS0901; SPASE_1; UNKNOWN_1.
DR PROSITE; PS0158; 2F_CCHC; 1.
FT CHAIN 1 524 Gag protein.
FT CHAIN 525 1759 Pol protein.
FT CHAIN 1719 2378 Env protein.
SQ SEQUENCE 2378 AA; 267306 MW; 6EC8DAEC1B4ED18A CRC64;

Alignment Scores:
Pred. No.: 1,31e-161 Length: 2378
Score: 2705.00 Matches: 497
Percent Similarity: 97.9% Conservative: 16
Best Local Similarity: 94.8% Mismatches: 11
Query Match: 95.0% Indels: 0
DB: 2 Gaps: 0

US-10-723-552-3_copy_585_2156 (1-1572) x Q9Q1X4_9GAMR (1-2378)

QY 1 ATGGGACAGCGGTGACGACCCCTCTTAGTTGACTCTCGACCATTTGCACTGAAGTTAAA 60
Db |||:::|||||
QY 1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20
Db |||:::|||||
QY 61 TCCAGGGCTCATATTTGTTCAGTTTCAGGTTAAGGAGGACCTTGGCAGACTTTCTGTGTC 120
Db |||:::|||||
QY 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysAla 40
Db |||:::|||||
QY 121 TCTGAATGGCGGACATTCGATTTGGATGGCCATCAGAGGGGACCTTTAAATCTCGAGATT 180
Db |||:::|||||
QY 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60
Db |||:::|||||
QY 181 ATCTGGCTGTTAAAGCAGTTATTTTTCAGACTCGACCCGGCTCTCATCCCGATCAGGAG 240
Db |||:::|||||
QY 61 IleuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80
Db |||:::|||||
QY 241 CCCTATATCTCTTACGTGGCAGATTGTCGAGAGGATCTCCGCCATGGTGTAAACCATGG 300
Db |||:::|||||
QY 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100
Db |||:::|||||
QY 301 CTGAATAGCCAAAGACCCAGGTCCCGAATTCCTGGCTCTTGGAGAGAAAAACAACAC 360
Db |||:::|||||
QY 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120
Db |||:::|||||
QY 361 TCGGCTGAAAAGTCACAGCCCTCTCTCATATCTACCCCGAGATTGAGGAGCCACCGGCT 420
Db |||:::|||||
QY 121 SerAlaGluLysValGluProSerProArgIleTyrProGluIleGluGluProProThr 140
Db |||:::|||||
QY 421 TGGCCGGAACCCCAATCTGTTTCCCCACCCCTTATCTGGCACAGGGTCCCGCGAGGGA 480
Db |||:::|||||
QY 141 TrpProGluProGlnProValProProProTyrProAlaGlnGlyAlaValArgGly 160
Db |||:::|||||
QY 481 CCCTTTGCCCTCTCTGGAGCTCCGGCGGTGAGGAGCACTGCTGCAGGAGACTCGGAGCCGG 540
Db |||:::|||||
QY 161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180
Db |||:::|||||
QY 541 AGGGGGCCACCCCGGAGCGGACAGAGATCCGACATTCACGCTCCGACAGTACCGC 600
Db |||:::|||||
QY 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly 200
Db |||:::|||||
QY 601 CCTCCACACCCGGGGGCCAATTGTCAGCCCTCCAGATTTGGCCCTTTTCTTCTGCAGAT 660
Db |||:::|||||
QY 201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220
Db |||:::|||||
QY 661 CTCATATATTTGGAATACTAACCATCCCTCTTCGGAGGATCCCAACGGCTCAGCGGG 720
Db |||:::|||||
QY 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240
Db |||:::|||||
QY 721 TTGGTGGAGTCCCTTATGTTCTCTCACCAGCCTACTTCGGATGATTGTCAACAGCTGCTG 780
Db |||:::|||||
QY 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260
Db |||:::|||||
QY 781 CAGACACTCTTTCAACAACCGGAGGAGAGAAATTTCTATTAGAGGCTAGAAAAATGTT 840
Db |||:::|||||
```

Db 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgGlyAsnVal 280  
Qy 841 CTGGGGCGGACGGGCGACCCACCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA 900  
Db 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
Qy 901 ACTGGCCCGGTTGGGACTACACACGGCTCAAGGTAGGAGAGCTTGAATCTATCGC 960  
Db 301 ThrArgProGlyTyrPhePyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
Qy 961 CAGGCTCTGGTGGGGGCTCCGGGGCCCTCAAGACGGCCCACTAATTTGGCTTAAGGTA 1020  
Db 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgProThrAsnLeuAlaLysVal 340  
Qy 1021 AGAAGTGTATGACGGGACCGAATGAACCCCTCTGTTTCTTGAGAGGCTCTGGAA 1080  
Db 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
Qy 1081 GCCTTCAGGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAGCCCTCAGTGGCT 1140  
Db 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
Qy 1141 TTGGCTTTTATAGACAGTACGCTTGTGATATTAGAAGAGCTTCAGAGCTCGAAGG 1200  
Db 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
Qy 1201 TTACAGGAGGCTGAGTTACGTGATCTAGTGAAGAGGCGAGAGAAGTATATTACAAAAGG 1260  
Db 401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg 420  
Qy 1261 GAGACAGAGAAAGAGGAAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAA 1320  
Db 421 GluThrGluGluGluLysGluGlnArgLysGluLysGluArgGluGluArgGluGluArg 440  
Qy 1321 CGTAATAACCGCAAGAGAGAAATTCACCTAAGATCTTGGCTGCAGTGGTTGAAGGAAA 1380  
Db 441 ArgAspArgGlnGluLysAsnLeuThrLysIleLeuAlaValAlaGluGlyLys 460  
Qy 1381 AGCAATACGAAAGAGAGAGAGATTTTAGGAAATAGGTCAGGCCCTAGACAGTCAGGG 1440  
Db 461 SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
Qy 1441 AACCTGGGCAATAGGACCCACTCGACAGGACCAATGTGCTATTTGTAAGAAAGAGGA 1500  
Db 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500  
Qy 1501 CACTGGGCAAGGACTGCCCCAGAGGGAACAAAGGACCAAGGACCAAGGATCTAGCTAGAA 1560  
Db 501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520  
Qy 1561 GAAGATAAAGAT 1572  
Db 521 GluAspLysAsp 524

RESULT 15  
O62708.PIG  
ID O62708.PIG PRELIMINARY; PRT; 524 AA.  
AC O62708;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag protein.  
OS Sus scrofa (pig).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
OC Sus.  
ON NCBI\_TaxID=9823;  
RX [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=minature swine.  
RX MEDLINE=98216827; PubMed=9557749;  
RA Akiyoshi D.E., Denaro M., Zhu H., Greenstein J.L., Banerjee P.,  
RA Fishman J.A.;

"Identification of a full-length cDNA for an endogenous retrovirus of miniature swine.";  
J. Virol. 72:4503-4507(1998).  
DR EMBL; AF038601; AAC16766.1; -; mRNA.  
DR HSSP; O9WTP4; 1B44.  
DR SMR; O62708; 7-98.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR003036; Gag\_p30.  
DR InterPro; IPR000840; G:retro matrix.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF011140; Gag\_MA; 1.  
DR Pfam; PF02093; Gag\_p30; 1.  
DR Pfam; PF00098; zf-CCHC; 1.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR SMART; SM00343; Znf\_C2HC; 1.  
DR PROSITE; PS0158; ZF\_CCHC; 1.  
SQ SEQUENCE 524 AA; 59577 MW; BEE1P250D7FAD18E CRC64;

## Alignment Scores:

Pred. No.: 1.73e-161 Length: 524  
Score: 2702.00 Matches: 496  
Percent Similarity: 97.9% Conservative: 17  
Best Local Similarity: 94.7% Mismatches: 11  
Query Match: 94.9% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x O62708\_PIG (1-524)

Qy 1 ATGGGACAGACGGTGCACGCCCTCTTAGTTTGAATCTCGACCATTCGACTGAAGTTAAA 60  
Db 1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20  
Qy 61 TCAGGGCTCATATTTGTCAGTTCAGTTAAGAGGACCTTGCGCAGACTTCTGTGTC 120  
Db 21 SerArgAlaHisAsnLeuSerValGlnValLysGlyProTrpGlnThrPheCysAla 40  
Qy 121 TCTGAATGGCCGACATTCGATGTTGGATGGCCATCAGAGGGACCTTTAATCTGAGATT 180  
Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
Qy 181 ATCTGGCTTTAAAGCAGTTATTTTTCAGACTGGACCCCGCTCTCATCCGATCAGAG 240  
Db 61 IleLeuAlaValLysAlaIleIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
Qy 241 CCTATATCTTACGTGGCAAGATTGGCAGAGGATCTCCCGCATGGGTTAAACCATGG 300  
Db 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100  
Qy 301 CTGAATAAGCCAAAGAACCCAGTCCCGGATTCGGCTCTTGAGAGAGAAAACAAACAC 360  
Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
Qy 361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCGAGATTGAGGAGCCACCGCT 420  
Db 121 SerAlaGluLysValGluProSerProArgIleTyrProGluIleGluGluProThr 140  
Qy 421 TGGCCGGAACCCCAATCTGTTCGCCCAACCCCTTATCTGGCAGAGGTCGCCGAGGGA 480  
Db 141 TrpProGluProGlnProValProProProProProProProProProProProProPro 160  
Qy 481 CCCTTTGCCCTCTCGAGCTCCGGCGGTGGAGGACCTCTCGAGGACTCGAGGCGG 540  
Db 161 ProSerAlaProProGlyAlaProValGluGlyProAlaAlaGlyThrArgSerArg 180  
Qy 541 AGGGGGCCACCCCGGAGCGGACAGACGAGATCGCACATTACCGCTGCGCAGTACGGC 600  
Db 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly 200  
Qy 601 CTTCCACACCGGGGGGCGCAATTGACGCCCTCCAGTATTGGCCCTTTCTCTCGAGAT 660

Db	201	ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp	220
Qy	661	CTCTATAAATGGAAAACTAAACCATCCCCCTTTCTCGGAGGATCCCCAACGCCTCAGGGG	720
Db	221	LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly	240
Qy	721	TTGTGTGAGTCCCTTATGTCTCTCACAGCCTACTTTGGGATGATTGTCAACAGCTGCTG	780
Db	241	LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu	260
Qy	781	CAGACACTCTTCAACCGAGGACGAGAGAGAACTCTATTAGAGGCTAGAAAAATGTT	840
Db	261	GlnThrLeuPheThrThrGluGluArgLeuGluLeuLeuGlnAlaLysLysAsnVal	280
Qy	841	CCTGGGCGCCAGCGGCACCCACCGGTGTGCAAAATGAGATTGACATGGGATTTCCCTTA	900
Db	281	ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu	300
Qy	901	ACTCGCCCGGTTGGGACTCAACACCGCTGAAGGTAGGGAGAGCTTGAAAAATCTATCGC	960
Db	301	ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg	320
Qy	961	CAGGCTCTGTTGGGGGTCTCCGGGGCGCTCAAGACGGCCCACTAAATTTGGCTCAAGGTA	1020
Db	321	GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal	340
Qy	1021	AGAGAAGTGATGACGGGACCGAATGAACCCCCCTCTCTTTTCTGTGAGAGGCTCTTGGAA	1080
Db	341	ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu	360
Qy	1081	GCCTTCAGGCGGTACACCCCTTTTGATCCCACTCAGAGGCCCAAAAGGCTCAGTGGCT	1140
Db	361	AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla	380
Qy	1141	TTGSCCTTTATAGCACAGTCAAGCCTTCGATATTAGAAAGAGCTTCAGAGACTGGAAGGG	1200
Db	381	LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly	400
Qy	1201	TTACAGGAGGCTGAGTTACGTGATCTAGTGAAGAGGCGAGAGAAAGTATATTACAAAGG	1260
Db	401	LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg	420
Qy	1261	GAGACAGNAGNAGAAAGGACAAAGAAAGAGAGAGAAAGAGAGNAAAGGAGGAAGA	1320
Db	421	GluThrGluGluGluLysGluGlnArgLysGluLysGluArgGluGluArgGluGluArg	440
Qy	1321	CGTAAATAAACCGCAAGAGAGAAATTCACATAAGATCTTGGCTGCAGTGGTTGAAGGAAA	1380
Db	441	ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys	460
Qy	1381	AGCAATACGGNAGAGAGAGAGATTTTAGGAAATATTAGTTCAGGCCCTAGACAGTCAAGG	1440
Db	461	SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly	480
Qy	1441	AACCTGGGCAATAGGACCCCACTCGACAGGACCAATGTCATATTTCTAAGAAAGAGGA	1500
Db	481	AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly	500
Qy	1501	CACCTGGGCAGGAACTGCCCAAGAGGGGAAAACAAAGACCAAGGATCTCTAGCTCTAGAA	1560
Db	501	HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGlu	520
Qy	1561	GAAGATAAAGAT	1572
Db	521	GluAspLysAsp	524

Search completed: February 14, 2006, 14:45:47  
Job time : 372.026 secs

•



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 12:52:14 ; Search time 41.8246 Seconds  
(without alignments)  
3302.856 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_585\_2156

Perfect score: 2848

Sequence: 1 ATGGAGCAGCGTGACGAC.....CTCTAGAGAGATAAAGAT 1572

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB.spool/US10723552/runat 14022006 125139 12833/app query.fasta\_1  
-DB=A Geneseq -QFMT=fastan -SUPFIX=rag -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US10723552 @CGN 1 1 734 @runat 14022006 125139 12833 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DRV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2817	98.9	524	2 AAW32096	Miniature
2	2817	98.9	524	4 AAB73285	Retoviral
3	2817	98.9	524	5 AAO19011	Porcine e
4	2817	98.9	524	8 ADS73446	Swine ret
5	2813	98.8	744	8 ADS73450	Swine ret
6	2705	95.0	524	5 AAO19009	Porcine e
7	2705	95.0	524	5 AAO19010	Porcine e
8	2702	94.9	524	2 AAW32091	Porcine r
9	2702	94.9	524	4 AAB73282	Defective

10	2702	94.9	524	8 ADS73441	AdS73441 Swine ret
11	2640	92.7	524	2 AAW32096	AAW32096 Porcine r
12	2640	92.7	524	4 AAB70629	AAB70629 Porcine e
13	1733	60.8	522	2 AAW81571	AAW81571 Mus dunni
14	1733	60.8	622	2 AAW81570	AAW81570 Mus dunni
15	1380.5	48.5	538	3 AAB10043	MLV gag
16	1380.5	48.5	538	7 ADH76469	ADH76469 Murine le
17	1380.5	48.5	648	2 RAY17946	RAY17946 MMLV gag
18	1380.5	48.5	1737	3 AAB10044	AAW10044 MMLV gag
19	1379.5	48.4	1784	2 AAR05898	AAR05898 Gene prod
20	1377.5	48.4	1784	2 AAR94427	AAR94427 FelV F6A
21	1374	48.2	580	2 AAW68402	AAW68402 FelV-A ga
22	1371	48.1	537	4 AAB49467	AAB49467 Canine re
23	1371	48.1	625	4 AAB49466	AAB49466 Canine re
24	1359	47.7	508	8 ADO03718	ADO03718 Feline le
25	1357	47.6	537	2 AAR75188	AAR75188 Oestroudu
26	1348.5	47.3	509	8 ADO03720	ADO03720 Feline le
27	930.5	32.7	499	5 ABB79874	ABB79874 Splen ne
28	842	29.6	441	7 ADH77072	ADH77072 Friend mu
29	842	29.6	529	7 ADH77071	ADH77071 Friend mu
30	832.5	29.2	253	4 AAM50120	AAM50120 Feline le
31	826	29.0	263	2 AAR80533	AAR80533 Moloney m
32	824	28.9	263	2 AAR80537	AAR80537 Moloney m
33	820	28.8	263	2 AAR80535	AAR80535 Moloney m
34	668	23.5	523	8 ADW50749	ADW50749 Human pro
35	553	19.4	405	4 ABG07968	ABG07968 Novel hum
36	553	19.4	430	4 ABG08938	ABG08938 Novel hum
37	553	19.4	1577	4 ABG06475	ABG06475 Novel hum
38	553	19.4	1577	4 ABG27123	ABG27123 Novel hum
39	553	19.4	1650	4 AAM39205	AAM39205 Human pol
40	543.5	19.1	611	4 AAM50122	AAM50122 Feline le
41	542.5	19.0	878	4 ABG06484	ABG06484 Novel hum
42	526.5	18.5	688	4 ABG08930	ABG08930 Novel hum
43	510.5	17.9	205	3 AAB39242	AAB39242 Human sec
44	478	16.8	505	4 ABG15313	ABG15313 Novel hum
45	475.5	16.7	208	8 ABO59520	ABO59520 Human gen

ALIGNMENTS

RESULT 1

AAW32096

ID AAW32096 standard; protein; 524 AA.

XX

AC AAW32096;

XX

DT 27-AUG-2003 (revised)

DT 09-FEB-1998 (first entry)

XX

DE Miniature swine retrovirus GAG protein.

XX

KW Retrovirus; porcine; GAG protein; xenotransplantation; infectious;  
KW provirus; organ transplant; donor; activated virus; PCR.

XX

OS Pig endogenous retrovirus.

XX

PH Key Location/Qualifiers

FT Protein 1..524

FT /label= GAG\_protein

XX

XX WO9721836-A1.

XX

XX 19-JUN-1997.

XX

PF 13-DEC-1996; 96WO-US019680.

XX

PR 14-DEC-1995; 95US-00572645.

XX

PA (GEO ) GEN HOSPITAL CORP.

XX

PI Fishman JA;

XX

DR WPI; 1997-332804/30.

DR N-PSDB; AAT74884.  
 XX New nucleic acid from porcine retro:viruses - used for detecting viruses  
 PT in transplant or other tissue and for assessing risk of transmitting  
 PT infection to graft recipient.  
 XX  
 PS Claim 22; Fig 3; 128pp; English.

XX This is a porcine retrovirus from miniature swine containing the coding  
 CC region for a putative viral GAG protein. This sequence and PCR fragments  
 CC generated from the sequence (see AAT74812-T74882) could be used to screen  
 CC organs for porcine retroviruses prior to xenotransplantation.  
 CC Transplantation can increase the likelihood of retroviral activation if  
 CC intact and infectious proviruses are present. The porcine retroviral  
 CC sequence can be used to generate probes to determine the level (e.g. copy  
 CC number) of intact (i.e. potentially replicating) porcine provirus  
 CC sequences in a strain of xenograft transplantation donors. It can be used  
 CC to detect mutations, genetic lesions or viral recombinants and also to  
 CC determine the histological localisation of activated retrovirus. Using  
 CC Polymerase Chain Reaction DNA Quantitation (PDQ) on blood mononuclear  
 CC cells, infectivity titration and susceptibility testing can be performed.  
 CC Ultimately animal donors without intact porcine retroviral sequences or a  
 CC lower copy number of viral elements could be selected. (Updated on 27-AUG  
 CC -2003 to correct OS field.)

SQ Sequence 524 AA;

#### Alignment Scores:

Pred. No.: 4.09e-241 Length: 524  
 Score: 2817.00 Matches: 524  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 98.9% Indels: 0  
 DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x AAW32096 (1-524)

QY 1 ATGGGACAGAGGTGACGACCCCTTCTAGTTTGACTCTCGACCATTCGAGTTAA 60  
 DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluVallys 20  
 QY 61 TCCAGGGCTCATAATTCTCGAGTTGAGTGAAGGACCTTGGCAGACTTTCTGTGC 120  
 DB 21 SerArgAlaHisAsnLeuSerValGlnVallylsysGlyProTrpGlnThrPheCysVal 40  
 QY 121 TCTGAATGCCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCGAGATT 180  
 DB 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluile 60  
 QY 181 ATCTGGCTGTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 240  
 DB 61 IleLeuAlaValLysAlaValIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
 QY 241 CCCTATATCTTACCTGGCAAGATTGGCAGAGGATCTCCGCCATGGGTTAAACATGG 300  
 DB 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProProTrpVallylsProTrp 100  
 QY 301 CTGAATAAGCCAAAGAACCCAGGTCCTCCGAATTCGGCTCTTGGAGAGAAAAACAAAC 360  
 DB 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
 QY 361 TCGGCTGAAAGTCAGCCCTCTCTCATATCTTACCCCGAGNTTGGAGGCCACCGGCT 420  
 DB 121 SerAlaGluLysValLysProSerProHisIleTrpProGluIleGluGluProProAla 140  
 QY 421 TGGCCGGAACCCCAATCTGTTCCCCACCCCTTATCTGGCAGACGGGTGCGCGAGGGGA 480  
 DB 141 TrpProGluProGlnSerValProProProTrpProTrpLeuAlaGlnGlnAlaalaArgGly 160  
 QY 481 CCCTTTGCCCTCTCTGGAGCTCCGGCGGTGAGGGACCTGTGCGAGGACCTCGGAGCCGG 540  
 DB 161 ProPheAlaProProGlyAlaProAlaValGluGlyProAlaAlaGlyThrArgSerArg 180

QY 541 AGGGGCGCCACCCCGGAGCGGACAGACGATCGGACATTACCGCTCGCGACGTCACGGC 600  
 DB 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyrGly 200  
 QY 601 CCTCCACACCGGGGGGCCAATTGCAGCCCTCCAGTATTGGCCCTTTTCTTCGCAAT 660  
 DB 201 ProProThrProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
 QY 661 CTCTATATTTGGAATCTACCATCCCTTCTCGGAGGATCCCAACGCTCACCAGG 720  
 DB 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
 QY 721 TTGTGTGAGTCCCTTATGTTCTCTCACGAGCTACTTGGGATGATTGTCAACAGCTGCTG 780  
 DB 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260  
 QY 781 CAGACACTCTTTCACACCGGAGCGAGAGAGAAATCTTATTAGAGGCTAGAAAAATGTT 840  
 DB 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280  
 QY 841 CCTGGGCGGACGCGGCGACCCACCGGTTGCCAAATGAGATTGACATGGGATTTCCCTTA 900  
 DB 281 ProGlyAlaAspGlyArgProThrArgLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
 QY 901 ACTCGCCCGGTTGGGACTACACACGCTGAAGCTAGGAGAGCTTGAAATCTATCGC 960  
 DB 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
 QY 961 CAGGCTCTGTGTGGCGGTCTCCGGGCGCTTCAAGCGGCCCTAAATTTGGCTAAGGTA 1020  
 DB 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
 QY 1021 AGAGAAGTGTCAGGACCGCAATGAACCCCTCTGTTTCTTGAGAGCTCTTCGAA 1080  
 DB 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuGlu 360  
 QY 1081 GCCTTCAGGGCGGTACACCCCTTTTATCCCACTCAGAGGCCCAAAAGCCCTCAGTGCT 1140  
 DB 361 AlaPheArgArgTyrThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
 QY 1141 TTGCGCTTTATPAGACAGTCAGCTTGGATATTAGAAGAAAGCTTCAGAGACTGGAAGG 1200  
 DB 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
 QY 1201 TTACAGAGGCTGATGTTACGTGATCTAGTAGGAGCGGAGAGAAAGTATATTACAAAAG 1260  
 DB 401 LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg 420  
 QY 1261 GAGACAGAGAAAGAAAGGAAACAAAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 DB 421 GluThrGluGluGluArgGluGlnArgLysGluArgGluArgGluArgGluArg 440  
 QY 1321 CGTAATAACCGCAGAGAGAAATTTGACTAAGATCTTGGCTGCGAGTGGTTGAAGGAAA 1380  
 DB 441 ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys 460  
 QY 1381 AGCAATACGGAACAGAGAGAGATTTTAGGAAAATTAGGTCAGGCCCTTAGACAGTCAGG 1440  
 DB 461 SerAsnThrGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
 QY 1441 AACCTGGGCAATAGGACCCCACTCGACAAGACCAATGTGCATATTGTAAAGAAAGAGGA 1500  
 DB 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluArgGly 500  
 QY 1501 CACTGGGCAAGGAATCTCCCAAGAGAGGAAACAAAGGACCAAGGATCCTAGCTCTAGAA 1560  
 DB 501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProArgIleLeuAlaLeuGlu 520  
 QY 1561 GAAGATAAGAT 1572  
 DB 521 GluAspLysAsp 524

RESULT 2

## AAB73285

ID AAB73285 standard; protein; 524 AA.

XX AAB73285;

XX AC AAB73285;

XX DT 23-MAY-2001 (first entry)

XX DE Retroviral protein #1 found in miniature swine.

XX XX

XX KW Retrovirus; graft transplantation; xenotransplantation; miniature swine.

XX OS Unidentified.

XX XX

XX PN US6190861-B1.

XX PD 20-FEB-2001.

XX XX

XX PF 13-DEC-1996; 96US-00766528.

XX XX

XX PR 14-DEC-1995; 95US-00572645.

XX XX

XX PA (GEO ) GEN HOSPITAL CORP.

XX PI Fishman JA;

XX XX

XX DR WPI; 2001-256211/26.

XX DR N-PSDB; AAF77727.

XX PT Assessing risk of endogenous retroviruses in clinical practice and in

XX PT xenotransplantation, comprises using probe sequences derived from swine

XX PT or miniature swine retroviral genome.

XX XX

XX PS Disclosure; Fig 3; 127pp; English.

XX XX

XX CC The present invention relates to a method for screening a cell or tissue

XX CC for the presence or expression of a retrovirus (RV), comprising

XX CC contacting a target nucleic acid from the cell or tissue with a second

XX CC nucleic acid from the present invention (e.g. AAF77727 or a fragment

XX CC thereof). The method is useful for RV detection and to assess graft

XX CC transplantation risk. Screening of animals allows the elimination of

XX CC donors with active replication of known viruses. Inactive proviruses can

XX CC be detected and inactivated, allowing identification and elimination of

XX CC potential human pathogens derived from swine in a manner not possible in

XX CC the outbred human organ donor population and is important to the

XX CC development of human xenotransplantation

XX XX

XX SQ Sequence 524 AA;

## Alignment Scores:

Pred. No.:	4.09e-241	Length:	524
Score:	2817.00	Matches:	524
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	98.9%	Indels:	0
DB:	4	Gaps:	0

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x AAB73285 (1-524)

QY	1	ATGGGACAGCGGTGACGACCCCTTGTAGTTGACTCTCGACCATTCGACTGAAGTTAAA	60
DB	1	MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluVal	20
QY	61	TCCAGGGCTCAATAATTTGTCAAGTTTCAAGTAAAGGACCTTCGACAGACTTCTGTGTC	120
DB	21	SerArgAlaHisAsnLeuSerValGlnValIysGlyProTrpGlnThrPheCysVal	40
QY	121	TCTGAATGGCCGACATTCGATGTTGGATGCGCATCAGAGGGGACCTTAAATTCGAGATT	180
DB	41	SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle	60
QY	181	ATCTCGCTGTTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCGATCAGAG	240
DB	61	IleLeuAlaValIysAlaValIlePheGlnThrGlyProGlySerHisProAspGlnGlu	80

QY	241	CCCTATATCTTACGTGGCAAGATTGGCAGAGGATCTCCGCCATGGTTTAAACCATGG	300
DB	81	ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValIysProTrp	100
QY	301	CTGAATAAGCCCAAGAACCCAGGTCCCGGAATTCGGCTCTGGAGAGAAAAACAAC	360
DB	101	LeuAsnIysProArgIysProGlyProArgIleLeuAlaLeuGlyGluIysAsnIysHis	120
QY	361	TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTTACCCCGAGATTGAGGAGCCACCGCT	420
DB	121	SerAlaGluIysValIysProSerProHisIleTrpGluIleGluGluProProAla	140
QY	421	TGGCCGGAACCCCAATCTGTTCCCCCACTTATCTGGCAGAGGTGCGCCGAGGGGA	480
DB	141	TrpProGluProGlnSerValProProProTrpIleuAlaGlnGlyAlaAlaArgGly	160
QY	481	CCCTTGGCCCTCTCTGGAGTCCCGCGGTGGAGGACCTGTGTCAGGGACTCGGAGCCGG	540
DB	161	ProPheAlaProProGlyAlaProAlaValGluGlyProAlaAlaGlyThrArgSerArg	180
QY	541	AGGGGCGCCACCCCGGAGCGGACAGACGAGATCCGACATTACCGCTCGCAGCTACCGC	600
DB	181	ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyrGly	200
QY	601	CCTCCACACCGGGGGCCCAATTGTCAGCCCTCCAGTATTGGCCCTTTTCTTCTGCAGAT	660
DB	201	ProProThrProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp	220
QY	661	CTCTATATTTGGAAAACTAACCATCCCTCTTTCGGAGGATCCCAACCGCTCACGGGG	720
DB	221	LeuTyrAsnTrpIysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly	240
QY	721	TTGGTGAGTCCCTTATGTTCTCCACAGCCTACTTGGGATGATTGTCAACAGCTGCTG	780
DB	241	LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu	260
QY	781	CAGACACTTTCACAAACCGAGGCGGAGAGAGATTTCTATTAGAGGCTAGAAAAAATGTT	840
DB	261	GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal	280
QY	841	CCTGGGGCCGACGGGCCACCCAGCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA	900
DB	281	ProGlyAlaAspGlyArgProThrArgLeuGlnAsnGluIleAspMetGlyPheProLeu	300
QY	901	ACTCGCCCGGTTGGGACTACACAGGCTCAGGTAGGAGGAGCTTGAATAATCTATCGC	960
DB	301	ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuIleTyrArg	320
QY	961	CAGGCTCTGCTGGCGGCTCTCCGGGGCGCTCAAGAGCGGCCCACTAATTTGGCTTAAGTA	1020
DB	321	GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaIysVal	340
QY	1021	AGAGAAATGATGACGGGACCCGAATGAACCCCTCTCTTTTCTTTGAGAGCTCTTGAA	1080
DB	341	ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuGlu	360
QY	1081	GCCTTCAGGCGGTACACCCCTTTTGTATCCACCTCAGAGGCCCAAAAGCTCAGTGGCT	1140
DB	361	AlaPheArgArgTyrThrProPheAspProThrSerGluAlaGlnIysAlaSerValAla	380
QY	1141	TTGGCTCTTATAGGACAGTCCAGCTTCGATATTAGAAAGAGCTTCAGAGACTCGAAGGG	1200
DB	381	LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgIysIysLeuGlnArgLeuGluGly	400
QY	1201	TTACAGAGGCTGAGTTTACCTGATCTAGTCAAGAGGACAGAGAAAGTATTATTACAAAGG	1260
DB	401	LeuGlnGluAlaGluLeuArgAspLeuValIysGluAlaGluIysValTyrTyrLysArg	420
QY	1261	GAGCAGACAGAAAGGGGACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1320
DB	421	GluThrGluGluGluArgGluGlnArgLysGluArgGluArgGluArgGluArgGluArg	440



QY 1081 GCCTTCAGCGGTACACCCCTTTTGTATCCCACTCAGAGGCCCAAAAAGCCTCAGTGCT 1140  
Dd |||||||  
361 AlaPheArgTyThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
QY 1141 TTGGCCTTTATAGGACAGTCAGCCTTGGATATAGAAAGAGCTTCAGAGACTGAAGG 1200  
Dd |||||||  
381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLeuGlnArgLeuGluGly 400  
QY 1201 TTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGAGGAGAGAAAAGTATATTACAAAAGG 1260  
Dd |||||||  
401 LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg 420  
QY 1261 GAGACAGAGAAGAAAGGGAACAAAGAAAGAGAGAGAAAGAGGAAAGGAGGAAAGA 1320  
Dd |||||||  
421 GluThrGluGluGluArgGluGlnArgLysGluArgGluArgGluGluArg 440  
QY 1321 CGTAATTAACGCGACAGAGAGAAATTTGACTAAGATCTTGGCTGCAGTGTTCAGAGGAAA 1380  
Dd |||||||  
441 ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys 460  
QY 1381 AGCAATACGGAAGAGAGAGAGATTTTAGGAAAATTAGGTCAAGGCCCTAGACAGTCAGGG 1440  
Dd |||||||  
461 SerAsnThrGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
QY 1441 AACCTGGCAATAGGACCCCACTCGACAAGGACCAATGTGCATATTTGTAAGAAAGAGGA 1500  
Dd |||||||  
481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluArgGly 500  
QY 1501 CACTGGGCAAGAACTGCCCAAGAGGGAACAAAGACCAAGGATCCTAGCTTAGAA 1560  
Dd |||||||  
501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProArgIleLeuAlaLeuGlu 520  
QY 1561 GAAGATAAAGAT 1572  
Dd |||||||  
521 GluAspLysAsp 524

RESULT 4  
ADS73446

ID ADS73446 standard; protein; 524 AA.

XX ADS73446;

DT 16-DEC-2004 (first entry)

DE Swine retroviral gag protein.

XX Swine retroviral protein; immunosuppressive; gene therapy; gag protein.

OS Porcine endogenous retrovirus.

XX US2004185435-A1.

FN 23-SEP-2004.

PD 26-NOV-2003; 2003US-00723552.

PF 14-DEC-1995; 95US-00572645.

PR 13-DEC-1996; 96US-00766528.

PR 14-SEP-2000; 2000US-00861858.

XX (GEHO ) GEN HOSPITAL CORP.

FA Fishman JA;

XX WPI; 2004-689179/67.

DR N-ESDB; ADS73369.

XX New porcine retroviral polypeptide encoded by a nucleic acid, useful in  
PT evaluating an immunosuppressive treatment for the ability to activate a  
PT retrovirus, such as an endogenous porcine retrovirus.

XX Claim 8; Fig 3; 83pp; English.

PS

XX The present invention relates to the swine retroviral polypeptides and  
CC their encoding nucleic acids. The methods and compositions of the present  
CC invention are useful for screening a cell or tissue, e.g. a heart, lung,  
CC liver, bone marrow, kidney, brain cells, neural tissue, pancreas and  
CC intestinal tissue xenograft, for the presence or expression of a swine or  
CC miniature swine retrovirus or retroviral sequence. The invention is also  
CC useful in evaluating an immunosuppressive treatment for the ability to  
CC activate a retrovirus such as an endogenous porcine retrovirus. The  
CC invention is also useful in gene therapy. The present sequence is the  
CC swine retroviral gag protein.

XX Sequence 524 AA;

Alignment Scores:

Pred. No.: 4.09e-241 Length: 524  
Score: 2817.00 Matches: 524  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 98.9% Indels: 0  
DB: 8 Gaps: 0

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x ADS73446 (1-524)

QY 1 ATGGGACAGACGGTGACGACCCCTCTTAGTTTGACTCTCGACCATTTGGACTGAAGTTAAA 60  
Dd |||||||  
1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuAspHisTrpThrGluValys 20  
QY 61 TCCAGGGCTCATATTTGTCTAGTTTCAGGTTAAGAGGACCTTGGCAGACTTCTGTGTC 120  
Dd |||||||  
21 SerArgAlaHisAsnLeuSerValGlnValLysGlyProTrpGlnThrPheCysVal 40  
QY 121 TCTCAATGGCGCACATTCGATTTGGATGGCATTGAGAGGGGACCTTTAATTTCTGAGATT 180  
Dd |||||||  
41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
QY 181 ATCTGGCTGTAAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCGATCAGGAG 240  
Dd |||||||  
61 IleLeuAlaValLysAlaValIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
QY 241 CCCTATATCTTAGTGGCAAGATTTCGACAGAGGATCTCGCCATGGGTTAAACCATGG 300  
Dd |||||||  
81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100  
QY 301 CTGAATAAGCCAAAGAACCCAGGTCGCCGAATTCGTCTCTTGAGAGAGAAAAACAAACAC 360  
Dd |||||||  
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
QY 361 TCGGCTCAAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGAGCCACCGGCT 420  
Dd |||||||  
121 SerAlaGluLysValLysProSerProHisIleTyrProGluIleGluGluProAla 140  
QY 421 TGGCCGGAACCCCAATCTGTTCCTCCCAACCCCTTATCTGGCAGAGGTCGCCGAGGGA 480  
Dd |||||||  
141 TrpProGluProGlnSerValProProProTyrLeuAlaGlnGlyAlaAlaArgGly 160  
QY 481 CCCTTTGCCCTCTCTCGAGCTCCGGCGGTGGAGGACCTCTGTCAGGAGCTCGCAGGCGG 540  
Dd |||||||  
161 ProPheAlaProGlyAlaProAlaValGluGlyProAlaAlaGlyThrArgSerArg 180  
QY 541 AGGGGCGCCACCCCGGAGCGGACAGAGATCGCAGCATTTACCGCTCGCAGCATACGGC 600  
Dd |||||||  
181 ArgGlyAlaThrProGluArgThrAspGluIleAlaThrLeuProLeuArgThrTyrGly 200  
QY 601 CCTCCGACACCGGGGGCCAAATTGACGCCCTCTCCAGTATTGGCCCTTTTCTTCTGCAGAT 660  
Dd |||||||  
201 ProProThrProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
QY 661 CTCTATAATTGGAAACTAACCATCCCTTTCTCGGAGGATCCCAACGCCCTCACCGGG 720  
Dd |||||||  
221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
QY 721 TTGGTGGAGTCCCTTATGTTTCTCTACACGACTACTTGGGATGATTGTCAACAGCTGCTG 780

```
|||||
241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeu 260
QY 781 CAGACACTTTCACAAACGAGGAGCGAGAGAAATTCATTAGAGCTAGAAAAATGTT 840
Db 261 GlnThrLeuPheThrThrGluGluArgGluArgGluLeuLeuGluAlaArgLysAsnVal 280
QY 841 CCTGGGCGGACGGGACCCACCGGTTGCMAATGAGATTGACATGGGATTCCTTA 900
Db 281 ProGlyAlaAspGlyArgProThrArgLeuGlnAsnGluIleAspMetGlyPheProLeu 300
QY 901 ACTGCCCCCGTGGACACTACAAACGCTGAAGCTAGCGAGCTTGAAATCTATCGC 960
Db 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320
QY 961 CAGGCTCTGGTGGCGGCTTCGGGGCGCCTCAAGACGGCCCACTAAATTTGGCTAAGTA 1020
Db 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340
QY 1021 AGAAGAGTGTGCAGGGACCGAATCAACCCCTCTGTTTCTTGAGAGGCTCTTGAA 1080
Db 341 ArgGluValMetGlnGlyProAsnGluProSerValPheLeuGluArgLeuGlu 360
QY 1081 GCCTTCAGCGGTACACCCCTTTTCATCCACTCAGAGGCCCAAAAGCCTCAGTGCT 1140
Db 361 AlaPheArgArgTyrThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380
QY 1141 TTGGCCTTTATAGACAGTCAGCTTGATATAGAAAGAGCTTCAGAGCTGGAAGG 1200
Db 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGly 400
QY 1201 TTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGCGAGAAAGTATATACAAAGG 1260
Db 401 LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg 420
QY 1261 GAGCAGAGAAGAAAGGAAACAAAGAAAGAGAGAGAAAGAGAGAAAGGAGGAAAGA 1320
Db 421 GluThrGluGluGluArgGluGlnArgLysGluArgGluArgGluGluArg 440
QY 1321 CGTAATAACCGCAAGAGAAATTTGACTTAAGATCTTTGGCTGCGAGTGGTTGAAGGAAA 1380
Db 441 ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValValGluGlyLys 460
QY 1381 AGCATACGGAAGAGAGAGATTTTAGGAAATTAGCTCAGGCCCTAGACAGTCAGGG 1440
Db 461 SerAsnThrGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480
QY 1441 AACCTGGGCAATAGGACCCCACTCGACAGGACCAATGTCATATTGTAAAGAAAGAGGA 1500
Db 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluArgGly 500
QY 1501 CACTGGGCAAGGAATGCCCCCAAGAGGAAACAAAGGACCAAGGATCTAGCTTAGAA 1560
Db 501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProArgIleLeuAlaLeuGlu 520
QY 1561 GAAGATAAAGAT 1572
Db 521 GluAspLysAsp 524
RESULT 5
ID ADS73450 standard; protein; 744 AA..
XX AC ADS73450;
XX DT 16-DEC-2004 (first entry)
XX DE Swine retroviral (Tsukuba-1) gag protein.
XX KW Swine retroviral protein; immunosuppressive; gene therapy; gag protein.
XX OS Porcine endogenous retrovirus.
XX
```

```

FH Key Location/Qualifiers
FT Misc-difference 53 /note= "Encoded by TGA"
FT Misc-difference 134 /note= "Encoded by TGA"
FT Misc-difference 134 /note= "Encoded by TGA"
XX US2004185435-A1.
PN 23-SEP-2004.
XX 26-NOV-2003; 2003US-00723552.
XX 14-DEC-1995; 95US-00572645.
XX 13-DEC-1996; 96US-00766528.
XX 14-SEP-2000; 2000US-00661858.
XX (GEHO ) GEN HOSPITAL CORP.
XX Fishman JA;
XX WPI; 2004-689179/67.
XX N-PSDB; ADS73367.
XX New porcine retroviral polypeptide encoded by a nucleic acid, useful in evaluating an immunosuppressive treatment for the ability to activate a retrovirus, such as an endogenous porcine retrovirus.
XX Claim 8; Fig 1; 83pp; English.
XX The present invention relates to the swine retroviral polypeptides and their encoding nucleic acids. The methods and compositions of the present invention are useful for screening a cell or tissue, e.g. a heart, lung, liver, bone marrow, kidney, brain cells, neural tissue, pancreas and intestinal tissue xenograft, for the presence or expression of a swine or miniature swine retrovirus or retroviral sequence. The invention is also useful in evaluating an immunosuppressive treatment for the ability to activate a retrovirus such as an endogenous porcine retrovirus. The invention is also useful in gene therapy. The present sequence is the swine retroviral (tsukuba-1) gag protein.
SQ Sequence 744 AA;
Alignment Scores:
Pred. No.: 1,03e-240 Length: 744
Score: 2813.00 Matches: 522
Percent Similarity: 100.0% Conservative: 2
Best Local Similarity: 99.6% Mismatches: 0
Query Match: 98.8% Indels: 0
DB: Gaps: 0
US-10-723-552-3_COPY_585_2156 (1-1572) x ADS73450 (1-744)
QY 1 ATGGGACAGACGGTGCAGACCCCTCTTAGTTTGACTCTCGACCATTTGACATGAAAGTTAAA 60
Db 221 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValLys 240
QY 61 TCCAGGGCTCATTAATTTGTCTCAGTTTAAAGAGGACCTTGGCAGACCTTCTGTGTC 120
Db 241 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysVal 260
QY 121 TCTGAATGGCCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAAATTTCTGAGATT 180
Db 261 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluLeu 280
QY 181 ATCCTGGCTGTTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCGATCAGAG 240
Db 281 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 300
QY 241 CCCTATATCCTTACGTGCAAGATTGGCAGAGGATCTCCGCCATGGGTAAACCATGG 300
Db 301 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 320
QY 301 CTGAATAAGCAAGAAAGCCAGGTCCCCGAATTTCTGGCTCTTTGGAGAGAAAAACAACAC 360
```

Db 321 LeuAsnLysProArgLysProGlyProArgLysLeuAlaLeuGlyGluLysAsnLysHis 340  
QY TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGAGCCACCGGCT 420  
Db 341 SerAlaGluLysValLysProSerProHisLysLysProGluLysGluGluProProAla 360  
QY TGGCCGGAAACCCCAATCTGTTCCCCACCCCTTATCTGGCAACAGGGTCCGCGAGGGA 480  
Db 361 TrpProGluProGlnSerValProProProLysValProProProLysLysGlnGluAlaAlaArgGly 380  
QY CCCTTTGGCCCTCTCTGGAGCTCCGGCGTGGAGGACCTGTCAGGAGCTCGAGCCGG 540  
Db 381 ProPheAlaProProGlyAlaProAlaValGluGlyProSerAlaGlyThrArgSerArg 400  
QY AGGGCGCCACCCCGGAGCGACAGACGAGATCGGCATTACCGCTGCGCAGCTACGCGC 600  
Db 401 ArgGlyAlaThrProGluArgThrArgGluLeuAlaThrLeuProLeuArgThrTyroGly 420  
QY CCTCCACACCGGGGGCCAAATGCGAGCCCTCCAGTATTCGCCCTTTTCTTCGACAT 660  
Db 421 ProProThrProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 440  
QY CTCTATAATTGGAAACTAAACCATCCCTCTTCTCGAGGATCCCAACGGCTCACGGGG 720  
Db 441 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 460  
QY TTGCTGGAGTCCCTTATGTTCTCTCACCGAGCTACTTGGGATGATTGTCAACACGCTG 780  
Db 461 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspPheCysGlnGlnLeu 480  
QY 781 CAGACACTTTCAACACGAGGAGCGAGAGAAATCTATTAGAGGCTAGAAAATGTT 840  
Db 481 GlnThrLeuPheThrThrGluGluArgGluArgLysLeuGluAlaArgLysAsnVal 500  
QY 841 CCTGGGCGGAGCGGACCCACCGCTCAACGCTCAAGGTAGGAGCTTGAAAATCTATCGC 960  
Db 501 ProGlyAlaAspGlyArgProThrArgLeuGlnAsnGluIleAspMetGlyPheProLeu 520  
QY 901 ACTCGCCCGGTGGGACTCAACACGCTCAAGGTAGGAGCTTGAAAATCTATCGC 960  
Db 521 ThrArgProGlyTrpAspPheAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 540  
QY 961 CAGGCTCTGGTGGCGGTCTCCGGGCGCCTCAAGAGCGGCCACTAATTTGGCTTAAGTA 1020  
Db 541 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 560  
QY 1021 AGAAGTGTATGAGGAGCGAATGAACCCCTCTGTTTCTTGAGAGGCTCTTGAA 1080  
Db 561 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuGlu 580  
QY 1081 GCCTTCAGGCGGTACACCCCTTTTGATCCCACTCAGAGGCCCAAAAGCTCAGTGGCT 1140  
Db 581 AlaPheArgTrpThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 600  
QY 1141 TTGGCCCTTTATGAGACAGTCAAGCTCTGGATATTAGAAAGAGCTTCAGAGCTGGAAGG 1200  
Db 601 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 620  
QY 1201 TTACAGAGGCTGAGTTAGTGATCTAGTGAGGAGCGCAGAGAAAGTATTATTCAAAGG 1260  
Db 621 LeuGlnGluAlaGluLeuArgAspLeuValLysGluAlaGluLysValTyrTyrLysArg 640  
QY 1261 GAGACAGAAGAAGAAAGGAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db 641 GluThrGluGluGluArgGluGlnArgGluGluArgGluGluArgGluGluArgGlu 660  
QY 1321 CGTAATAAAGCGCAAGAGAGAAATTTGACTAAGATCTTGGCTGCGAGTGGTTGAAGGAAA 1380  
Db 661 ArgAsnLysArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGlyLys 680  
QY 1381 AGCAATACGGAAGAGAGAGATTTTAGGAAATTTAGGTCAGGCCCTCAGCAGTCAAGG 1440

Db 681 SerAsnThrGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 700  
QY 1441 AACCTGGCAATAGCAGCCCTCGACCAAGCACCATGTGTATATTGTAAAGAAAGAGA 1500  
Db 701 AsnLeuGlyAsnAlaThrProLeuAspLysAspGlnCysAlaTyrCysLysGluArgGly 720  
QY 1501 CACTGGCAGAGAACTCCCCCAAGAGGGAAACAAAGACCAAGGATCTTAGCTCTAGAA 1560  
Db 721 HistPalaArgAsnCysProLysLysGlyAsnLysGlyProArgIleLeuAlaLeuGlu 740  
QY 1561 GAAGATAAAGAT 1572  
Db 741 GluAspLysAsp 744  
RESULT 6  
ID AA019009 standard; protein; 524 AA.  
XX AA019009;  
AC AA019009;  
DT 07-NOV-2002 (first entry)  
XX Porcine endogenous retrovirus type A gag.  
DE PERV; protease; virucide; infection; xenotransplant; gag.  
XX Porcine endogenous retrovirus.  
OS DE10138528-Al.  
PN 04-JUL-2002.  
XX 06-AUG-2001; 2001DE-01038528.  
XX 23-DEC-2000; 2000DE-01064897.  
XX (UYMU-) UNIV MUECHEN MAXIMILIANS LUDWIG.  
XX Von Der Helm K, Blusch JH, Seelmeir S;  
DR WPI; 2002-609974/66.  
DR N-PSDB; AAL49333.  
XX New nucleic acid, useful for identifying specific inhibitors for  
controlling infection following xenotransplantation, encodes porcine  
endogenous retrovirus protease.  
XX Disclosure; Page 14-16; 30pp; German.  
XX The present invention relates to proteases and their coding sequences  
from porcine endogenous retroviruses (PERVs) type A, B and C. The coding  
sequences are useful in screening for inhibitors of proteases,  
particularly where retroviral. The inhibitors are potentially useful for  
treating or preventing PERV infections in humans, specifically those who  
have received a pig organ transplant. The present sequence is a  
retroviral gag protein  
XX Sequence 524 AA;  
Alignment Scores:  
Pred. No.: 3,68e-231 Length: 524  
Score: 2705.00 Matches: 497  
Percent Similarity: 97.9% Conservative: 16  
Best Local Similarity: 94.8% Mismatches: 11  
Query Match: 95.0% Indels: 0  
DB: Gaps: 5  
US-10-723-552-3\_COPY\_585\_2156 (1-1572) x AA019009 (1-524)  
QY 1 ATGGGACAGCGGTGAGGACCCCTCTAGTTTCTGACTCTCGACCTGGACTGAGTTAAA 60  
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20



QY 61 TCCAGGGCTCATAATTTCTCAGTTTAAAGAGGACCTTGGCAGACTTCTGTGTC 120  
Db |||||  
21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysAla 40  
QY 121 TCTGAATCGCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAAATTCAGATT 180  
Db |||||  
41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
QY 181 ATCTGGCTGTTAAAGCAGTTATTTTTCAGACTGGACCCGGCTCTCATCCGATCAGGAG 240  
Db |||||  
61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
QY 241 CCCTATATCCTTACGTGTCAGATTGTCAGAGGATCTCCGCCATGTTGGTAAACCATGG 300  
Db |||||  
81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100  
QY 301 CTGAATAAGCCAAGAACCCAGGTCCCGAATCTTGGCTCTTGGAGAGAAAACAAACAC 360  
Db |||||  
101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
QY 361 TCGGCTGAAAAGTCAGCCCTCTCTCATATCTACCCCGAGATTGAGGACCCAGCGCT 420  
Db |||||  
121 SerAlaGluLysValGlnProSerProArgIleTyrProGluIleGluGluProProThr 140  
QY 421 TGGCGGAACCCCAATCTGTTCCCGACCCCTTATCTGGCACAGGGTCCCGAGGGGA 480  
Db |||||  
141 TrpProGluProGlnProValProProProProTyrProAlaGlnGlyAlaValArgGly 160  
QY 481 CCTTTGTCCTCTCTGGAGCTCCGCGGTGGAGGACCTGTGTCAGGGACTCGGAGCCGG 540  
Db |||||  
161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180  
QY 541 AGGGCGCCACCCCGAGGAGACAGACGAGATCGCGACATTACCGCTCGGCACGTACGGC 600  
Db |||||  
181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly 200  
QY 601 CCTCCACACCGGGGGCCAAATGTCAGCCCTCCAGTATTCGCCCTTTCTTCGCGAT 660  
Db |||||  
201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
QY 661 CTCTATAATTGGAACCTAACCATCCCTTTCTCGAGAGATCCCAACGCTCTCACGGGG 720  
Db |||||  
221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
QY 721 TTGTGGAGTCCCTTATGTTCTCTCACCAGCTACTTGGGATGATGTCAACAGCTGCTG 780  
Db |||||  
241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260  
QY 781 CAGACACTCTTCACAAACGAGGAGCGAGAGAGATTCTATTAGAGGCTAGAAAAAATGTT 840  
Db |||||  
261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaArgLysAsnVal 280  
QY 841 CTGCGGCGCGAGCGGCGACCCACGCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA 900  
Db |||||  
281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
QY 901 ACTGCGCCCGGTGGGACTACAAACGCGCTGAAGCTAGGAGAGCTTGAAATCTATCGC 960  
Db |||||  
301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
QY 961 CAGGCTCTGTTGGCGGGTCTCGGGCGCTCAAGACGCCCACTTAATTTGCTTAAGGTA 1020  
Db |||||  
321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
QY 1021 AGAGAAGTGATGCAGGGACCGGAATGAACCCCTCTGTTTTTCTTGAGAGGCTCTTGAA 1080  
Db |||||  
341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
QY 1081 GCCTTCAGCGGTACACCCCTTTTGCATCCCACTCAGAGGCCCAAAAAGCTTCAGTGGCT 1140  
Db |||||  
361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
QY 1141 TTGGCCTTTATAGGACAGTCAAGCTTGGATATTAGAAAGAGCTTCAGAGACTGGAAGG 1200

Db |||||  
381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
QY 1201 TTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAGGAGAGAAAGTATATATACAAAAGG 1260  
Db |||||  
401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg 420  
QY 1261 GAGCACAGNAGAAAGGACAAAGAAAGAGAGAGAGAAAGAGAGAAAGGAGGAAAGA 1320  
Db |||||  
421 GluThrGluGluGluLysGluGlnArgLysGluLysGluArgGluGluGluArg 440  
QY 1321 CGTAATAAACCAGCAAGAGAAGAAATTCCTAAAGATCTTGGCTGCAGTGGTTGAAGGAAA 1380  
Db |||||  
441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaAlaValValGluGlyLys 460  
QY 1381 AGCAATACGGAAGAGAGAGAGATTTAGGAAAATTAGGTGAGCCCTAGACAGTCAGGG 1440  
Db |||||  
461 SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
QY 1441 AACCTGGCAATAGGACCCCACTCGACAAGACCAATGTCATATTGTAAGAAGAGGA 1500  
Db |||||  
481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500  
QY 1501 CACTGGCAGAACTGCCCAAGAGGAAACAAAGACCAAGGATCCTAGCTCTAGAA 1560  
Db |||||  
501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520  
QY 1561 GAAGATAAAGAT 1572  
Db |||||  
521 GluAspLysAsp 524  
RESULT 7  
AAO19010  
ID ARO19010 standard; protein; 524 AA.  
XX  
AC AAO19010;  
XX  
DT 07-NOV-2002 (first entry)  
XX  
DE Porcine endogenous retrovirus type B gag.  
XX  
KW PERV; protease; virucide; infection; xenotransplant; gag.  
XX  
OS Porcine endogenous retrovirus.  
XX  
FN DE10138528-Al.  
XX  
PD 04-JUL-2002.  
XX  
PF 06-AUG-2001; 2001DE-01038528.  
XX  
PR 23-DEC-2000; 2000DE-01064897.  
XX  
PA (UTMU-) UNIV MUENCHEN MAXIMILIANS LUDWIG.  
XX  
PI Von Der Helm K, Blusch JH, Seelmeir S;  
XX  
DR WPI; 2002-609974/66.  
XX  
PT N-PSDB; AAL49334.  
XX  
PT New nucleic acid, useful for identifying specific inhibitors for  
controlling infection following xenotransplantation, encodes porcine  
endogenous retrovirus protease.  
XX  
PS Disclosure; Page 16-17; 30pp; German.  
XX  
CC The present invention relates to proteases and their coding sequences  
from porcine endogenous retroviruses (PERVs) type A, B and C. The coding  
sequences are useful in screening for inhibitors of proteases, useful for  
particulary where retroviral. The inhibitors are potentially useful for  
treating or preventing PERV infections in humans, specifically those who  
have received a pig organ transplant. The present sequence is a  
retroviral gag protein



PF 13-DEC-1996; 96WO-US019680.  
 XX PR 14-DEC-1995; 95US-00572645.  
 XX XX (GBHO) GEN HOSPITAL CORP.  
 XX FI Fishman JA;  
 XX XX MPI; 1997-332804/30.  
 DR N-PSDB; AAT74883.  
 XX XX  
 PT New nucleic acid from porcine retroviruses - used for detecting viruses  
 PT in transplant or other tissue and for assessing risk of transmitting  
 PT infection to graft recipient.  
 XX  
 XX Claim 16; Fig 2; 128pp; English.  
 XX  
 CC This sequence represents a putative viral GAG protein isolated from a  
 CC defective porcine retrovirus found in PK-15 cells. This sequence and PCR  
 CC fragments generated from it (see AAT74812-T74882) could be used to screen  
 CC organs for porcine retroviruses prior to xenotransplantation.  
 CC Transplantation can increase the likelihood of retroviral activation if  
 CC intact and infectious proviruses are present. The porcine retroviral  
 CC sequence can be used to generate probes to determine the level (e.g. copy  
 CC number) of intact (i.e. potentially replicating) porcine provirus  
 CC sequences in a strain of xenograft transplantation donors. It can be used  
 CC to detect mutations, genetic lesions or viral recombinants and to  
 CC determine the histological localisation of activated retroviruses. Using  
 CC Polymerase Chain Reaction DNA Quantitation (PDQ) on blood mononuclear  
 CC cells, infectivity titration and susceptibility testing can be performed.  
 CC Ultimately animal donors without intact porcine retroviral sequences or  
 CC with a lower copy number of viral elements could be selected. (Updated on  
 CC 27-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 524 AA;

Alignment Scores:  
 Pred. No.: 6,81e-231 Length: 524  
 Score: 2702.00 Matches: 496  
 Percent Similarity: 97.9% Conservative: 17  
 Best Local Similarity: 94.7% Mismatches: 11  
 Query Match: 94.9% Indels: 0  
 DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x AAW32091 (1-524)

Qy 1 ATGGGACAGCGTGACGACCCCTCTTAGTTGACTCTCGACCAATGGACTGAAGTTAAA 60  
 Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20  
 Qy 61 TCCAGGGCTCATATTTGTGAGTTCAGGTTAAGAGGGACCTTGGCAGACTTCTGTGTC 120  
 Db 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysAla 40  
 Qy 121 TCTGAATGGCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAAATCTGAGATT 180  
 Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60  
 Qy 181 ATCTGGCTGTAAAGCAGATTATTTTTCAGACTGGACCCGGCTCTCATCCCGCATCAGGAG 240  
 Db 61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
 Qy 241 CCTATATCTTACGTGGCAGATTGGCAGAGATCTCCGCCNTGGGTTAAACCATGG 300  
 Db 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProProTrpValLysProTrp 100  
 Qy 301 CTGAATAGCCAAAGAACAGGTCCTCCGAATCTGGCTCTTGGAGAGAAAAACAACAC 360  
 Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
 Qy 361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGACCCCGGCT 420  
 Db 121 SerAlaGluLysValGluProSerProArgIleTyrProGluIleGluGluProProThr 140

Qy 421 TGGCCGGAAACCCCAATCTGTTCCCCACCCCTTATCTGGCAGAGGGTGC CGCGAGGGA 480  
 Db 141 TrpProGluProGlnProValProProProProProProAlaGlnGlyAlaValArgGly 160  
 Qy 481 CCCTTTGCCCTCTCTGGAGCTCCGGCGTGGAGGACCTGCTGAGGAGACTCGAGCGG 540  
 Db 161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180  
 Qy 541 AGGGCGCCACCCCGGAGCGGACAGAGATCCGACATTTACCGCTCGCGCACGTACGGC 600  
 Db 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly 200  
 Qy 601 CCTCCACACACGGGGGCCAATTTCAGCCCTCCAGTATTGGCCCTTTCTTCTGCAGAT 660  
 Db 201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220  
 Qy 661 CTCTATAATTGGAAAACTAACCATCCCTTTCTCGGAGGATCCCAACGCTCACGGGG 720  
 Db 221 LeuTyrAsnTrpLysThrAsnHisProPheSerGluAspProGlnArgLeuThrGly 240  
 Qy 721 TTGTGGAGTCCCTTATGTTCTCTCACAGCTACTTGGGATGATTTGTCACAGCTGCTG 780  
 Db 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260  
 Qy 781 CAGACACTCTTCACACCGGAGGCGAGAGAGAAATCTTATTAGAGGCTAGAAAAATGTT 840  
 Db 261 GlnThrLeuPheThrThrGluGluArgGluArgLeuLeuGluAlaLysLysAsnVal 280  
 Qy 841 CCTGGGGCCGACGGGCGACCCACGCGTTCGCAAAATGAGATTGATGGGATTTCCCTTA 900  
 Db 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
 Qy 901 ACTGCCCGGTTGGGACTACACACGCTGAAGTAGGAGGAGCTTGAAAAATCTATCGC 960  
 Db 301 ThrArgProGlyTrpAspTyrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrArg 320  
 Qy 961 CAGGCTCTGTGGCGGTCTCCGGGCGCCTCAAGACGGCCCTAATTTGGCTAAAGTA 1020  
 Db 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
 Qy 1021 AGAGAAGTGTATGACAGGACCGAATGAACCCCTCTCTGTTTTCTTGAGAGCTCTTGAA 1080  
 Db 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
 Qy 1081 GCCTTCAGGGGTATACCCCTTTTGATCCACCTCAGAGCCCAAAAAGCCTCAGTGGCT 1140  
 Db 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
 Qy 1141 TTGGCTTTATAGACAGTCTAGTCTGATATTAGAAGAGCTTCAGAGACTCGAAGGG 1200  
 Db 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGly 400  
 Qy 1201 TTACAGAGGCTGAGTTACGTGATCTAGTGAAGAGGAGCAGAGAAAGTATATTACAAAAG 1260  
 Db 401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrTyrArgArg 420  
 Qy 1261 GAGCAGAGAAGAAAGGGAACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 1320  
 Db 421 GluThrGluGluGluLysGluGlnArgLysGluLysGluArgGluGluArgGluArg 440  
 Qy 1321 CGTAATAACGGCAAGAGAGAGAAATTTGACTAAGATCTTGGCTGAGTGGTTGAAGGGAAA 1380  
 Db 441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaIleValAlaGluGlyLys 460  
 Qy 1381 AGCAATACGGAAGAGAGAGAGATTTTAGGAAAAATTAGGTCAGGCCCTCTAGACGTACGG 1440  
 Db 461 SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480  
 Qy 1441 AACTGGGCAATAGGACCCCTCGACAGGACCAATGTGCATATTGTGAAGAAGAGAGGA 1500  
 Db 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500

QY 1501 CACTGGCAAGAACTGCCCAAGAGCGGAACAAAGACCAAGGATCCTAGCTCTAGAA 1560  
|||||  
Db 501 HisTrpAlaArgAsnCyseProLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520  
|||||  
QY 1561 GAAAGATAAGAT 1572  
|||||  
Db 521 GluAspLysAsp 524  
|||||  
RESULT 9  
AAB73282  
ID AAB73282 standard; protein; 524 AA.  
XX  
AC AAB73282;  
XX  
DT 23-MAY-2001 (first entry)  
XX  
DE Defective retroviral genome protein #1 isolated from PK-15 cell line.  
XX  
KW Retrovirus; graft transplantation; xenotransplantation; PK-15 cell line.  
XX  
OS Unidentified.  
XX  
PN US6190861-B1.  
XX  
PD 20-FEB-2001.  
XX  
PF 13-DEC-1996; 96US-00766528.  
XX  
PR 14-DEC-1995; 95US-00572645.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
XX  
PI Fishman JA;  
XX  
DR WPI; 2001-256211/26.  
DR N-PSDB; AAF77726.  
XX  
PT Assessing risk of endogenous retroviruses in clinical practice and in  
PT xenotransplantation, comprises using probe sequences derived from swine  
PT or miniature swine retroviral genome.  
XX  
PS Disclosure; Fig 2; 127pp; English.  
XX  
CC The present invention relates to a method for screening a cell or tissue  
CC for the presence or expression of a retrovirus (RV), comprising  
CC contacting a target nucleic acid from the cell or tissue with a second  
CC nucleic acid from the present invention (e.g. AAF77726 or a fragment  
CC thereof). The method is useful for RV detection and to assess graft  
CC transplantation risk. Screening of animals allows the elimination of  
CC donors with active replication of known viruses. Inactive proviruses can  
CC be detected and inactivated, allowing identification and elimination of  
CC potential human pathogens derived from swine in a manner not possible in  
CC the outbred human organ donor population and is important to the  
CC development of human xenotransplantation  
XX  
SQ Sequence 524 AA;  
Alignment Scores:  
Pred. No.: 6.81e-231 Length: 524  
Score: 2702.00 Matches: 496  
Percent Similarity: 97.9% Conservative: 17  
Best Local Similarity: 94.7% Mismatches: 11  
Query Match: 94.9% Indels: 0  
DB: 4 Gaps: 0  
US-10-723-552-3\_COPY\_585\_2156 (1-1572) x AAB73282 (1-524)  
QY 1 ATGGGACAGACGGTACGACACCCCTCTAGTTTGAATCTCGACCAATTGGACTGAAGTTAAA 60  
|||||  
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20  
|||||  
QY 61 TCCAGGCTCATAAATTTGTCAAGTTCAAGTAAAGGACCTTGGCAGACTTCTGTGTC 120  
|||||

Db 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysAla 40  
|||||  
QY 121 TCTGAATGGCCGACATTCGATGTTGGATGGCCATCAGAGGGAGACCTTTAAATCTCAGATT 180  
|||||  
Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluLeu 60  
|||||  
QY 181 ATCTGGCTGTTAAAGCAGTTATTTTCAGACTGGACCCCGCTCTCATCCGATCAGAG 240  
|||||  
Db 61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80  
|||||  
QY 241 CCTATATCTTACGTGGCAAGATTTCGCAGAGGATCCTCCGCATCGGTAAACCATCGG 300  
|||||  
Db 81 ProTrpIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100  
|||||  
QY 301 CTGAATAAGCCAAAGACCCAGGTCCCGCAATTCGTGCTCTGGAGAGAAAAACACAC 360  
|||||  
Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120  
|||||  
QY 361 TCGCTGAAAAAGTCAAGCCCTCTCTCATATATCTACCCCGAGATGAGGAGCCACCGCT 420  
|||||  
Db 121 SerAlaGluLysValGluProSerProArgIleTrpProGluIleGluGluProProThr 140  
|||||  
QY 421 TGGCCGGAACCCCAATCTGTTCCGCCACCCCTTATCTGGCACAGGTCGCCGAGGGA 480  
|||||  
Db 141 TrpProGluProGlnProValProProProTrpProAlaGlnGlyAlaValArgGly 160  
|||||  
QY 481 CCTTTTGGCTCTCTGGAGCTCCCGCGGTGGAGGACCTGTCTGAGGAGCTCGGAGCCGG 540  
|||||  
Db 161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180  
|||||  
QY 541 AGGGCGCCACCCCGAGCGGACAGACGAGATCCGACATTCACGCTGCGCACGTACGGC 600  
|||||  
Db 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrgly 200  
|||||  
QY 601 CCTCCACACCGGGGGGCAATTCCAGCCCTCCAGTATTCGCCCTTTCTCTTCGCAGAT 660  
|||||  
Db 201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTrpProPheSerSerAlaAsp 220  
|||||  
QY 661 CTCTATAATTTGAAAACTAAACATCCCTTTTCGGAGGATCCCAACGCTCACCGGG 720  
|||||  
Db 221 LeuTyrsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240  
|||||  
QY 721 TTGCTGGAGTCCCTTATGTTCTCTCACAGCTACTTTGGGATGATTTGTCAACAGCTGCTG 780  
|||||  
Db 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnLeuLeu 260  
|||||  
QY 781 CAGACATCTTCAACACCGAGCGGAGAGAGAAATCTATTAGAGCTAGAAAAATGTT 840  
|||||  
Db 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuGluAlaLysLysVal 280  
|||||  
QY 841 CCTGGGCGCACGGGCGACCCAGCGGTTCCAAATGAGATTGACATGGGATTTCCCTTA 900  
|||||  
Db 281 ProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluIleAspMetGlyPheProLeu 300  
|||||  
QY 901 ACTCGCCCGGTTGGGACTACAACACCGCTGAAGGTAGGAGAGCTTGAATAATCTATCGC 960  
|||||  
Db 301 ThrArgProGlyTrpAspTrpAsnThrAlaGluGlyArgGluSerLeuLysIleTyArg 320  
|||||  
QY 961 CAGGCTCTGTGGCGGTCTCCGGGCGCTCAAGACGGCCCACTAATTTGGCTTAAGTA 1020  
|||||  
Db 321 GlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVal 340  
|||||  
QY 1021 AGAGAGTGTATGAGGAGCGCAATGACCCCTCTGTTTCTTGTAGAGGCTCTTGGAA 1080  
|||||  
Db 341 ArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGlu 360  
|||||  
QY 1081 GCCTTCAGGCGGTACACCCCTTTTGATCCCACTCCAGAGGCCCAAAAAAGCCTCAGTGGCT 1140  
|||||  
Db 361 AlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAla 380  
|||||  
QY 1141 TTGGCTTTATAGACAGTCAAGCTTGGATATTAGAAAGAGCTTCAGAGACTGGAAGG 1200  
|||||  
Db 381 LeuAlaPheIleGlyGlnSerAlaLeuAspIleArgLysLeuGlnArgLeuGluGly 400  
|||||

QY 1201 TTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGAGGAGAGAAAGTATATTACAAAAGG 1260  
Db 401 LeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrArgArg 420

QY 1261 GAGACAGAAAGAAAGGAAACAAGAAAGAGAGAGAAAGAGAGAAAGGAGGAAAGA 1320  
Db 421 GluThrGluGluGluLysGluArgLysGluLysGluArgGluGluGluArg 440

QY 1321 CGTAATAACCGCAAGAGAGAAATTTGCTAAGATCTTGGCTGACGTGTTGAAGGGAAA 1380  
Db 441 ArgAspArgArgGlnGluLysAsnLeuThrLysIleLeuAlaValValGluGluLys 460

QY 1381 AGCAATACGGAAGAGAGAGAGATTTTAGGAAAATTTAGTCCAGGCCCTPAGACAGTCAGGG 1440  
Db 461 SerSerArgGluArgGluArgAspPheArgLysIleArgSerGlyProArgGlnSerGly 480

QY 1441 AACCTGGGCAATAGGACCCCACTCGACAAGGACCAATGTGCATATTTGTAAGAAGAGGA 1500  
Db 481 AsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrCysLysGluLysGly 500

QY 1501 CACTGGGCAAGAACTGCCCAAGAGGAAACAAGGAAACAAGGAACTTAGCTCTAGAA 1560  
Db 501 HisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGlu 520

QY 1561 GAAGATAAAGAT 1572  
Db 521 GluAspLysAsp 524

RESULT 10  
AD573441  
ID AD573441 standard; protein; 524 AA.  
XX AD573441;  
XX 16-DEC-2004 (first entry)  
DT Swine retroviral gag protein fragment.  
DE Swine retroviral protein; immunosuppressive; gene therapy; gag protein.  
KW Porcine endogenous retrovirus.  
OS US2004185435-A1.  
XX 23-SEP-2004.  
XX 26-NOV-2003; 2003US-00723552.  
XX 14-DEC-1995; 95US-00572645.  
PR 13-DEC-1996; 96US-00766528.  
PR 14-SEP-2000; 2000US-00661858.  
XX (GEHO ) GEN HOSPITAL CORP.  
PA Fishman JA;  
XX WPI; 2004-689179/67.  
DR N-PSDB; AD573368.  
XX New porcine retroviral polypeptide encoded by a nucleic acid, useful in  
PT evaluating an immunosuppressive treatment for the ability to activate a  
PT retrovirus, such as an endogenous porcine retrovirus.  
XX Claim 8; Fig 2; 83pp; English.  
XX The present invention relates to the swine retroviral polypeptides and  
CC their encoding nucleic acids. The methods and compositions of the present  
CC invention are useful for screening a cell or tissue, e.g. a heart, lung,  
CC liver, bone marrow, kidney, brain cells, neural tissue, pancreas and  
CC intestinal tissue xenograft, for the presence or expression of a swine or  
CC miniature swine retrovirus or retroviral sequence. The invention is also  
CC useful in evaluating an immunosuppressive treatment for the ability to

CC activate a retrovirus such as an endogenous porcine retrovirus. The  
CC invention is also useful in gene therapy. The present sequence is the  
CC swine retroviral gag protein fragment.

XX Sequence 524 AA;

Alignment Scores: Pred. No.: 6,81e-231 Length: 524  
Score: 2702.00 Matches: 496  
Percent Similarity: 97.9% Conservativity: 17  
Best Local Similarity: 94.7% Mismatches: 11  
Query Match: 94.9% Indels: 0  
DB: Gaps: 0

US-10-723-552-3\_copy\_585\_2156 (1-1572) x AD573441 (1-524)

QY 1 ATGGGACAGACGGTGAAGACCCCTTGTAGTTTGACTCTCGACCAATTGGACTGAAGTTAAA 60  
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20

QY 61 TCCAGGGCTCATATTTGTCTAGTTTCAGTTTCAAGTAAAGAGGACCTTGGCAGACTTCTGTGTC 120  
Db 21 SerArgAlaHisAsnLeuSerValGlnValLysGlyProTrpGlnThrPheCysAla 40

QY 121 TCTGAATGGCCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCGAGATT 180  
Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60

QY 181 ATCTGGCTGTTAAAGCAGTTATTTTCAGACTGGACCCGGCTCTCATCCCGATCAGGAG 240  
Db 61 IleLeuAlaValLysAlaIlePheGlnThrGlyProGlySerHisProAspGlnGlu 80

QY 241 CCTATATCTTACGTGGCAAGATTTCGCAGAGATCTCCGCCATGGTTAAACCATGG 300  
Db 81 ProTyrIleLeuThrTrpGlnAspLeuAlaGluAspProProTrpValLysProTrp 100

QY 301 CTGAATAAGCCAGAAAGCCAGGTCCCGAATTCGGCTCTTGGAGAGAAAAACAAACAC 360  
Db 101 LeuAsnLysProArgLysProGlyProArgIleLeuAlaLeuGlyGluLysAsnLysHis 120

QY 361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGACCCCGGCT 420  
Db 121 SerAlaGluLysValGluProSerProArgIleTyrProGluIleGluGluProProThr 140

QY 421 TGGCGGNAACCCCAATCTGTTCCCCACCCCTTATCTGGCCACAGGGTCCCGCAGGGGA 480  
Db 141 TrpProGluProGlnProValProProProTrpProProProTrpProAlaGlnGlyAlaValArgGly 160

QY 481 CCTTTGCCCCCTCTGGAGCTCCCGCGTGGAGGGACCTGCTGCAGGGACTCGGAGCCGG 540  
Db 161 ProSerAlaProProGlyAlaProValValGluGlyProAlaAlaGlyThrArgSerArg 180

QY 541 AGGGCGCCACCCCGGAGCGGACAGACGAGATCCGACATTACCGCTGCGCACGTACGGC 600  
Db 181 ArgGlyAlaThrProGluArgThrAspGluIleAlaIleLeuProLeuArgThrTyrGly 200

QY 601 CCTCCACACCGGGGGCCCAATTCAGCCCTCCAGTATTCGCCCTTCTTCTTCGCAGAT 660  
Db 201 ProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrTrpProPheSerSerAlaAsp 220

QY 661 CTCTATTAATGGAAAACTAACCAATCCCTTCTCGGAGGATCCCCACGCTCACGGGG 720  
Db 221 LeuTyrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGly 240

QY 721 TTGTTGGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGGATGATTGTCAACAGCTGCTG 780  
Db 241 LeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeu 260

QY 781 CAGACACTCTTCACAAACCGAGGAGGAGAGATTCTTATTAGAGGCTAGAAAAATCTT 840  
Db 261 GlnThrLeuPheThrThrGluGluArgGluArgIleLeuLeuLeuAlaLysLysAsnVal 280

QY 841 CCTGGGGCCGACGGGCGACCCACCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTA 900



Db 180 gArgGlyAlaThrProGluuArgThrAspGluLeuAlaLeuProLeuArgThrTyrgl 200  
Qy 600 CCCTCCACACCGGGGGCCCAATTGCACCCCTCCAGTATTGGCCCTTTCTTTCGCAGA 659  
Db 200 yProProMetProGlyGlyGlnLeuGlnProLeuGlnTyrrpProPheSerSerAlaAs 220  
Qy 660 TCTCTAATATTGGAAACTACCATCCCTTTCGGAGGATCCCAACGCCCTCACGG 719  
Db 220 pLeuTyrrAsnTrpLysThrAsnHisProProPheSerGluAspProGlnArgLeuThrGl 240  
Qy 720 GTTGGTGAGTCCCTTATGTTCTCTCACCGACTACTTGGGATGATTGTCAACAGCTGCT 779  
Db 240 yLeuValGluSerLeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLe 260  
Qy 780 GCAGACACTTTCACAAACCGAGGAGCGAGAGAAATCTATTAGAGGCTAGAAAATGT 839  
Db 260 uGlnThrLeuPheThrThrGluuArgGluArgileLeuLeuGluAlaArgLysAsnVa 280  
Qy 840 TCCTGGGGCCGACGGGCGACCAACCGCTTGCAAAATGAGATTGACATGGGATTTCCCTT 899  
Db 280 lProGlyAlaAspGlyArgProThrGlnLeuGlnAsnGluileAspMetGlyPheProLe 300  
Qy 900 AACTCGCCCGCTTGGGACTCAACACCGCTGAAAGTAGGAGAGCTTGAATAATCTATCG 959  
Db 300 uThArgProGlyTrpAspTyrrAsnThrAlaGluGlyArgGluSerLeuLysIleTyrrAr 320  
Qy 960 CCAGCTCTGTGGCGGTCTCCGGGCGCTCAAGACGGCCCACTAATTTGGCTAAGGT 1019  
Db 320 gGlnAlaLeuValAlaGlyLeuArgGlyAlaSerArgArgProThrAsnLeuAlaLysVa 340  
Qy 1020 AAGAGAACTGATCGAGGACCGAATGAACCCCTCTCTTTTCTTGAGAGCTCTTGA 1079  
Db 340 lArgGluValMetGlnGlyProAsnGluProProSerValPheLeuGluArgLeuMetGl 360  
Qy 1080 AGCTTTCAGGCGGTACACCCCTTTTGATCCCACTTCAGAGGCGCCAAAAGCTCAGTGGC 1139  
Db 360 uAlaPheArgArgPheThrProPheAspProThrSerGluAlaGlnLysAlaSerValAl 380  
Qy 1140 TTTGGCTTTATAGACAGTCAAGCTTCGATATTAGAAAGAGCTTCAGAGCTGGAAGG 1199  
Db 380 aLeuAlaPheileGlyGlnSerAlaLeuAspIleArgLysLysLeuGlnArgLeuGluGl 400  
Qy 1200 GTTACAGAGCTGTGTTACGTGATCTAGTGAAGGAGGACAGAAAGTATATTACAAAG 1259  
Db 400 yLeuGlnGluAlaGluLeuArgAspLeuValArgGluAlaGluLysValTyrrArgAr 420  
Qy 1260 GGAGACAGAAGAAAGGAAACAAAGAAAGAGAGAGAAAGAGAGGAAAGGAGAAAG 1319  
Db 420 gGluThrGluGluLysGluGlnArgLysGluLysGluArgGluuArgGluuArg 440  
Qy 1320 ACCTAATAAACCGCAAGAGAAGAAATTTGACTAAGATCTTGGCTGCAGTGGTTGAAGGAA 1379  
Db 440 gArgAspArgArgGlnGluLysAsnLeuThrLysileLeuAlaAlaValValGluGlyLy 460  
Qy 1380 AAGCAATACGGAAGAGAGAGAGATTTAGGAAATTAGTTCAGGCCCTAGACAGTCAAG 1439  
Db 460 sSerSerArgGluuArgGluuArgAspPheArgLysileArgSerGlyProArgGlnSerGl 480  
Qy 1440 GAACCTGGGCAATAGGACCCCACTCGACAGGACCAATGTGCATATTGTAAAGAAAGAGG 1499  
Db 480 yAsnLeuGlyAsnArgThrProLeuAspLysAspGlnCysAlaTyrrCysLysGluLysGl 500  
Qy 1500 ACATGGGCAAGAACTGCCCCAAGAGGGGAAACAAAGGACCAAGGATCTTAGCTCTAGA 1559  
Db 500 yHisTrpAlaArgAsnCysProLysLysGlyAsnLysGlyProLysValLeuAlaLeuGl 520  
Qy 1560 AGAAGATAAAGAT 1572  
Db 520 uGluAspLysAsp 524

RESULT 12  
AAB70629

ID AAB70629 standard; protein; 524 AA.  
XX AAB70629;  
AC 15-MAY-2001 (first entry)  
XX Porcine endogenous retrovirus gag amino acid sequence.  
XX Porcine endogenous retrovirus; PoERV; gag; env; detection; antigenic;  
KW immunogenic; antiviral; vaccine; antiserum; viral infection.  
XX Porcine endogenous retrovirus.  
OS WO200112816-A2.  
XX 22-FEB-2001.  
XX 16-AUG-2000; 2000WO-GB003159.  
XX 18-AUG-1999; 99GB-00019604.  
XX (QUIP-) QUIP TECHNOLOGY LTD.  
XX Galbraith D, Kelly H, Smith K;  
XX WPI; 2001-211223/21.  
XX New porcine endogenous retrovirus polypeptide fragment with retroviral  
PT specific antigenic or immunogenic activity, for detection of retroviral  
PT antibodies in a sample and in therapy or diagnosis.  
XX Claim 3; Fig 1; 43pp; English.  
XX The present invention describes a porcine endogenous retrovirus (PoERV)  
CC polypeptide fragment (I) which has PoERV specific antigenic or  
CC immunogenic activity. Also described are: (1) an antiserum (II) specific  
CC to (I); (2) a PoERV specific antibody (III) or its fragment raised  
CC against (I); (3) use of (I) in detecting (III) in a sample; (4) use of  
CC (III) in detecting PoERV in a sample; (5) use of (I) or (III) in therapy  
CC or diagnosis; and (6) an assay kit (IV) comprising (I) or (III) for use  
CC in detection of PoERV in a sample. (I) has antiviral activity, and can be  
CC used in vaccine production. (I) is useful in the detection of PoERV  
CC antibodies (III) in a sample and (III) is useful in the detection of  
CC PoERV in a sample. (I) and (III) are useful in therapy or diagnosis.  
CC Antiserum (II) to (I) and kits (IV) comprising (I) or (III) are useful in  
CC detection of specific PoERV types. Viral infection can be monitored and  
CC by using (I) or (III). The present sequence represents the specifically  
CC claimed PoERV gag amino acid sequence from the present invention  
XX SQ Sequence 524 AA;  
Alignment Scores:  
Pred. No.: 2,21e-225 Length: 524  
Score: 2640.00 Matches: 495  
Percent Similarity: 97.3% Conservative: 16  
Best Local Similarity: 94.3% Mismatches: 13  
Query Match: 92.7% Indels: 2  
DB: 4 Gaps: 0  
US-10-723-552-3\_COPY\_585\_2156 (1-1572) x AAB70629 (1-524)  
Qy 1 ATGGGACAGACGGTGACGACCCCTCTTAGTTTGACTCTCGACCATTTGGACTGAAGTTAAA 60  
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpThrGluValArg 20  
Qy 61 TCCAGGGCTCATAAATTTGTTCAGTTTAAAGAGGGACCTTGGCAGACTTCTGTGTC 120  
Db 21 SerArgAlaHisAsnLeuSerValGlnValLysLysGlyProTrpGlnThrPheCysAla 40  
Qy 121 TCTGAATGGCCGACATTCGATGTTGGATGGCCATCAGAGGGACCTTAAATTCAGATT 180  
Db 41 SerGluTrpProThrPheAspValGlyTrpProSerGluGlyThrPheAsnSerGluIle 60





cell for producing a RDRV particle; (2) methods for producing a RDRV particle comprising a heterologous gene of interest; (3) cultured packaging cell line PD23; and (4) a RDRV produced by a method as in (2). The MDEV receptor is present on a variety of cells rendering MDEV pseudovirus packaging cells useful in methods of mammalian and particularly human gene transfer for gene therapy. The MDEV packaging cells are a stable and reproducible source of retroviral particles. Clones may be isolated from these populations that produce high titre virus. The packaging cell lines may be selected and cloned for other desirable properties, such as stability of in vivo growth, lack of production of helper virus, lack of reinfection by viral particles packaged in the cell, stability from genetic rearrangement and recombinational events, resistance to complement lysis, and improved ability to infect cells from higher mammals

XX Sequence 522 AA;

Alignment Scores:  
 Pred. No.: 9,08e-145 Length: 522  
 Score: 1733.00 Matches: 340  
 Percent Similarity: 73.3% Conservative: 61  
 Best Local Similarity: 62.2% Mismatches: 96  
 Query Match: 60.8% Indels: 50  
 DB: 2 Gaps: 12

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x AAW81571 (1-522)

QY 1 ATGGGACAGCGGTGACGACCCCTCTTAGTTGACTCTCGACCAATTGACATGAAGTTAAA 60  
 DB 1 MetGlyGlnSerValSerThrProLeuSerLeuThrLeuGluHisTrpLysGluVallys 20  
 QY 61 TCCAGGCTCATATTTCTAGTTTCAAGTTAAGAGGACCTTGGCAGACTTCTGTCTC 120  
 DB 21 IleArgAlaHisAsnGlnSerValGluValArgLysGlyProTrpGlnThrPheCysAla 40  
 QY 121 TCTGAATGCGCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTAAATTTCTGAGATT 180  
 DB 41 SerGluTrpProThrPheGlyValGlyTrpProProGluGlyAlaPheAspLeuSerLeu 60  
 QY 181 ATCTGGCTGTTAAGCAGTATTATTTTCAGACTGGACCCGCTCTCATCCCGATCAGAG 240  
 DB 61 IleAlaAlaValArgIleValPheGlnGlu---GluGlyGlyHisProAspGlnIle 79  
 QY 241 CCTATATCTTACGTGCAAGATTGTCAGAGGATCTCCGCCATGTCGGTTAAACCATGG 300  
 DB 80 ProTyrIleValThrTrpGlnSerLeuValGlnPheProProSerTrpValIysProTrp 99  
 QY 301 CTGAATAAGCAAGAAAGCCAGGTCCTCCGAATTCCTGGCTCTTGAGAGAAAAACAACAC 360  
 DB 100 -----ThrProAsnProSerLysLeuThrValAlaValAlaGlnSerAsp 114  
 QY 361 TCGGCTGAAAGTCAAGCCCTCT-----CCTCATATCTACCCCGAGATTGAG----- 408  
 DB 115 AlaAlaGluLysSerGlyProSerAlaProProLysIleTyrProGluIleAspAspLeu 134  
 QY 409 -----GAGCCACCGCTTGGCGGAAACCCCAATCTGTTCCC-----CCA 447  
 DB 135 LeuTrpMetAspSerGlnProProTrpProLeuProGlnGlnProProAlaAlaAla 154  
 QY 448 CCCCCTTATCTGGCAGCGGTGCCGCGAGGGACCCCTTGGCCCTCTCTGGAGCTCCGCG 507  
 DB 155 ProProValAlaAla-----ProGlnProGluProThrAlaSerGly 168  
 QY 508 GTGAGGACCTGCTGACGAGACTCGAGCCGGAGGGCGCCACCCCGAGCGG----- 561  
 DB 169 AlaGlnGlyProAlaGlyGlyThrArgSerArgArgGlyArgSerProAlaGluGluGly 188  
 QY 562 ---ACAGACGAGATCGGCACATTACCGCTGGCAGCTAC---GGCCCTCCACACCGGG 615  
 DB 189 GlyProAspSerThrValAlaLeuProLeuArgAlaHisValGlyGlyProThrProGly 208  
 QY 616 ---GGCCAATGTCAGCCCTCCAGTATTGGCCCTTTCTCTGTCAGATCTCTATAATTGG 672

DB 209 ProAsnAspLeuIleProLeuGlnTyrTrpProPheSerSerSerAspLeuTyrAsnTrp 228  
 QY 673 AAAACTAACCATCCCTTCTCGAGAGATCCCAACGCTCAGCGGGTGGTGGAGTCC 732  
 DB 229 LysThrAsnHisProProPheSerGluAsnProSerGlyLeuThrGlyLeuGluSer 248  
 QY 733 CTTATGTTCTCTCACAGCCTTACTTGGATGATTGTCAACAGCTGCTGCACACTCTTC 792  
 DB 249 LeuMetPheSerHisGlnProThrTrpAspAspCysGlnGlnLeuLeuGlnValPhe 268  
 QY 793 ACAACGAGGAGCGAGAGAGAAATCTATTAGAGCTAGAAAAATGTTCTCGGGCCGAC 852  
 DB 269 ThrThrGluGluArgGluArgIleLeuMetGluAlaArgLysAsnValLeuGlyGluAsp 288  
 QY 853 GGGGACCCAGCGGTTCGAAAAATGAGATTGACATGGATTTCCTTAACCTCGCCCGGT 912  
 DB 289 GlyThrProThrAlaLeuProAsnLeuValAspGluAlaPheProLeuAsnArgProAsn 308  
 QY 913 TGGGACTACACACGCTGAGGTAGGAGAGCTTGAATAATCTATCGCAGGCTCTGGTG 972  
 DB 309 TrpAspTyrAsnThrAlaGluGlyArgGlyArgLeuLeuValTyrArgTrpLeuVal 328  
 QY 973 GCGGCTCTCCGGGCGCTCAAGACGCGCCACTAATTTGGCTAAGTAAGAGAAGTATG 1032  
 DB 329 AlaGlyLeuArgGlyAlaAlaArgArgProThrAsnLeuAlaLysValArgGluValLeu 348  
 QY 1033 CAGGACCGAATGAACCCCTCTGTTTCTTTCAGAGGCTCTTGAAGCTTCAGGCGG 1092  
 DB 349 GlnGlyGlnThrGluProProSerValPheLeuGluArgLeuMetGluAlaTyrArgArg 368  
 QY 1093 TACACCCCTTTGATCCACCTCAGAGCGCCCAAAAGCTCAGTGGCTTTGGCCTTTATA 1152  
 DB 369 TyrThrProPheAspProSerSerGluGlyGlnLysAlaAlaValAlaMetAlaPheIle 388  
 QY 1153 GGACAGTCAGCTTGGATATTAGAAAGAAGCTTCAGAGACTGGAAGGTTACAGAGGCT 1212  
 DB 389 GlyGlnSerAlaProAspIleLysLysLysLeuGlnArgLeuGluGlyLeuGlnAspTyr 408  
 QY 1213 GAGTTAGTGTATCTAGTAGAGGCGAGAGAAAGTATTATTAACAAGGGACACAGAGAA 1272  
 DB 409 ThrLeuGlnAspLeuValLysGluAlaGluLysValTyrHisLysArgGluThrGluGlu 428  
 QY 1273 GAAAGGAAACAAAG 1332  
 DB 429 GluArgGlnGluArgGluLysLysGluValGluGluArgGluAsnArgArgAspArgArg 448  
 QY 1333 CAAGAGAAGATTTGACTAAGATCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1392  
 DB 449 GlnGluArgAsnLeuSerLysIleLeuAlaAlaValIleAsnAspArg----- 464  
 QY 1393 AGACAGAGAGATTTTAGAAAAATTAGTCAGGCCCTAGACAGTACAGGAACTCGGCAAT 1452  
 DB 465 -----GlnSerGluLysGlyArgTrpGlyPheLeuGlyAsn 476  
 QY 1453 AGG-----ACCCACTCGAAGAGCAATGTGCATAT 1485  
 DB 477 ArgAlaValLysProProGlyGlyArgLysThrProLeuGluLysAspGlnCysAlaPhe 496  
 QY 1486 TGTAAAGAAAGAGGACACTGGGCAAGGAACTGCCCCAAGAGAGGAAACAAAGGACCAAG 1545  
 DB 497 CysLysGluLysGlyHisTrpAlaLysAspCysProLysLys---ArgArgGlnPheLys 515  
 QY 1546 ATCTAGCTCTAGAAGAAGAT 1566  
 DB 516 ValLeuThrLeuGluAspAsp 522  
 RESULT 14  
 AAW81570  
 ID AAW81570 standard; protein; 622 AA.  
 XX AAW81570;  
 XX  
 DT 01-MAR-1999 (first entry)



```

QY 1033 CAGGACCGAATGAACCCCTCTGTTTTCTTGAGAGGCTCTTGGAGCCCTTCAGGGG 1092
Db 449 GlnGlyGlnThrGluProProSerValPheLeuGluArgLeuMetGluAlaTyrArgArg 468
QY 1093 TACACCCCTTTTGATCCCACTTCAGAGGCCCAAAAGCCTCAGTGGCTTTGGCTTTTATA 1152
Db 469 TyrThrProPheAspProSerSerGluGlyGlnLysAlaAlaValAlaMetAlaPheIle 488
QY 1153 GGACAGTCAGCTTGGATATTAGAAAGCTTCAGAGACTGGAGGTTTACAGAGGCT 1212
Db 489 GlyGlnSerAlaProAspIleLysLysLysLeuGlnArgLeuGluGlyLeuGlnAspTyr 508
QY 1213 GAGTTACTGTATAGTCAAGGAGCAGAGAAAGTATATTACAAAGGAGCAGAGAAGAA 1272
Db 509 ThrLeuGlnAspLeuValLysGluAlaGluLysValTyrHisLysArgGluThrGluGlu 528
QY 1273 GAAAGGGAACAAAGAAAGAGAGAAAGAGAGAAAGGAGAAAGCGTAATAAAGCG 1332
Db 529 GluArgGlnGluArgGluLysLysGluValGluGluArgGluAsnArgAspArgArg 548
QY 1333 CAAGAGAAGAAATTGACTTAAGATCTTGGCTGAGTGGTGAAGGAAAGCAATACGGA 1392
Db 549 GlnGluArgAsnLeuSerLysIleLeuAlaValIleAsnAspArg----- 564
QY 1393 AGAGAGAGAGATTTAGGAAATAGGTGAGGCCCTAGACAGTCAGGGAACCTGGCAAT 1452
Db 565 -----GlnSerGluLysGlyArgThrGlyPheLeuGlyAsn 576
QY 1453 AGG-----ACCCCACTCGACAGGCAACCAATGTGCATAT 1485
Db 577 ArgAlaValLysProProGlyGlyArgLysThrProLeuGluLysAspGlnCysAlaPhe 596
QY 1486 TGTTAAGAAAGAGACACTGGGCAAGGAACCTGCCCAAGAGAGGAAACAAAGGCCAAGG 1545
Db 597 CysLysGluLysGlyHisTrpAlaLysAspCysProLysLys---ArgArgGlnPheLys 615
QY 1546 ATCTAGCTCTAGAGAGAT 1566
Db 616 ValLeuThrLeuGluAspAsp 622

RESULT 15
AAB10043
ID AAB10043 standard; protein; 538 AA.
AC AAB10043;
XX
DT 02-NOV-2000 (first entry)
DE MMLV gag protein.
XX
KW Glycoprotein; gag gene; pol gene; GP-1; GP-2; anti-HIV; cytostatic;
KW gene therapy; treatment; infectious disease; HIV; AIDS; neoplasm;
KW carcinoma; melanoma.
XX
OS Moloney murine leukemia virus.
XX
FN EPI006196-A2.
XX
PD 07-JUN-2000.
XX
PF 25-NOV-1999; 99EP-00250415.
XX
PR 26-NOV-1998; 98DE-01056463.
XX
PA (PETT-) PETTE INST HEINRICH.
XX
PI Von Laer MD;
XX
DR WPI; 2000-378268/33.
XX
PT New retroviral packing cell useful as pharmaceutical carrier in gene
PT therapy for treatment of HIV and neoplasms, comprises retroviral genes
PT and glycoproteins.

```

```

XX
PS
XX
XX
CC This invention describes a novel retroviral packing cell (I), comprising
CC the retroviral genes gag, pol and glycoproteins GP-1 and GP-2 of the LCMV
CC coding gene gp, or a part of these. The products of the invention have
CC anti-HIV and cytostatic activity and can be used for gene therapy. (I) is
CC useful for in vitro infection of cells, especially hematopoietic stem
CC cells, for expression of transgenes in cells and as a pharmaceutical
CC carrier for gene therapy. (I) is therefore useful in the treatment of
CC infectious diseases (HIV or AIDS), neoplasms (carcinomas), melanomas and
CC other diseases. This sequence represents the Moloney murine leukemia
CC virus (MMLV) gag protein which is described in the method of the
CC invention
XX
SQ Sequence 538 AA;

```

Alignment Scores:

Pred. No.: 1,96e-113 Length: 538  
Score: 1380.50 Matches: 283  
Percent Similarity: 62.2% Conservative: 69  
Best Local Similarity: 50.0% Mismatches: 141  
Query Match: 48.5% Indels: 73  
DB: 3 Gaps: 13

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x AAB10043 (1-538)

```

QY 1 ATGGAGCAGAGCGGTGAGACCCCTTGTAGTTTGTGACTCTCGACCATTTGAGCTGAAGTTAAA 60
Db 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20
QY 61 TCCAGGGCTCATATTTGTCTAGTTTCAGTTTAAAGAGGACCTTCGCAGACTTCTGTCTC 120
Db 21 ArgIleAlaHisAsnGlnSerValAspValLysLysArgArgTrpValThrPheCysSer 40
QY 121 TCTGAATGGCCGACATTCGATTTGGATGGCCATCAGAGGGACCTTTAAATTCAGAGATT 180
Db 41 AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu 60
QY 181 ATCTGGCTGTAAAGCAGATTATTTTCAGACTCGACCCGGCTCTCATCCGATCAGAG 240
Db 61 IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal 80
QY 241 CCTATATCTTACGTGGCAGATTTCGAGAGATCTCCGCCATCTCCGCTGGTTAAACATGG 300
Db 81 ProTyrIleValThrTrpGluAlaPheAspProProProTrpValLysProPhe 100
QY 301 CTGAATAAGCCAAAGCCAGGTCCTCCGAATTCCTGCTCTTGGAGAGAAACAAACAC 360
Db 101 ValHis---ProLysProProProProProProSerAlaProSerLeuProLeuGlu 119
QY 361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGAGGCCA---CCG 417
Db 120 ProProArgSerThrProProArgSerSerLeuTyrProAlaLeuThrProSerLeuGly 139
QY 418 GCTTGGCCGGAACCCCA-----TCT 438
Db 140 AlalysProLysProGlnValLeuSerAspSerGlyGlyProLeuIleAspLeuLeuThr 159
QY 439 GTTCCCCCAACCCCTTATCTGCGACAGGTCCTCCGAGGAGCCCTTTCGCCCTCCT--- 495
Db 160 GluAspProProProTyr-----ArgAspProArgProProProSer 173
QY 496 -----GGAGCTCCGCGGTGGAGGACCT----- 519
Db 174 AspArgAspGlyAsnGlyGlyGluAlaThrProAlaGlyGluAlaProAspProSerPro 193
QY 520 ---GCTGCAGGACTCCGAGCCGAGGGGCCACCCCGAGCGGACACGAGATCGCG 576
Db 194 MetAlaSerArgLeuArgGlyArgGluProProValAlaAspSerThrThrSerGln 213
QY 577 ACATTACCGCTGCGCAGCTAGCGCCCTCCACACACCGGGGGCCCAATTGACGCCCTCCAG 636
Db 577 ACATTACCGCTGCGCAGCTAGCGCCCTCCACACACCGGGGGCCCAATTGACGCCCTCCAG 636

```

Disclosure; Page 31-32; 69pp; German.

```
Db      214 AlaPheProLeuArgAlaGlyGly-----AsnGlyGln-----LeuGln 226
Qy      637 TATTGGCCCTTTCTTCTGACAGATCTCTATAATTGGAATACTTAACCATCCCTTCTTCG 696
Db      227 TyrTrpProPheSerSerAspLeuTyrAsnTrpLysAsnAsnProSerPheSer 246
Qy      697 GAGNATCCCAACGCTCAGCGGGTGTGGAGTCCCTTATGTTCTCTCACCAGCTACT 756
Db      247 GluAspProGlyLysLeuThrAlaLeuIleGluSerValLeuIleThrHisGlnProThr 266
Qy      757 TGGGATGATTGTCAACAGCTGTCAGACACTCTTCAACCGGAGGCGAGAGAGATT 816
Db      267 TrpAspAspCysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGluGluLysGlnArgVal 286
Qy      817 CTATTAGAGGCTAGAAAAAATGTTCTCGGGCGGACCGGCGACCCACCGCGGTGCAAAAT 876
Db      287 LeuLeuGluAlaArgLysAlaValArgGlyAspAspGlyArgProThrGlnLeuProAsn 306
Qy      877 GAGATTGACATGGGATTTCCCTTAACCTGCCCGGTTGGGACTACACACGCTGAAGGT 936
Db      307 GluValAspAlaAlaPheProLeuGluArgProAspTrpAspTyrThrGlnAlaGly 326
Qy      937 AGGAGAGCTTGAATAATCTATCGCCAGGCTCTGGTGGCGGTCTCCGGGCGGCTCAAGA 996
Db      327 ArgAsnHisLeuValHisTyrArgGlnLeuLeuLeuAlaGlyLeuGlnAsnAlaGlyArg 346
Qy      997 CGGCCCACTAATTTGGCTAAGGTAAAGAGAGTGCAGAGGACCGAATGAACCCCTCT 1056
Db      347 SerProThrAsnLeuAlaLysValLysGlyIleThrGlnGlyProAsnGluSerProSer 366
Qy      1057 GTTTTCTTGAGAGCTCTTGAGAGCTTTCAGAGCTTTCAGCGGTACACCCCTTTGATCCACCTCA 1116
Db      367 AlaPheLeuGluArgLeuLysGluAlaTyrArgArgTyrThrProTyrAspProGluAsp 386
Qy      1117 GAGGCCCAAAAGCCTCAGTCGCTTTGGCTTTATAGACAGCTCAGCTTCGATATTAGA 1176
Db      387 ProGlyGlnGluThrAsnValSerMetSerPheIleTrpGlnSerAlaProAspIleGly 406
Qy      1177 AAGAAGCTTCAGAGACTGGAAGGTTTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAG 1236
Db      407 ArgLysLeuGluArgLeuGluAspLeuLysAsnLysThrLeuGlyAspLeuValArgGlu 426
Qy      1237 GCAGAGAAATATATTCAAAAGGAGACAGAGAAAGGAAACAAAGAAAGAGAGA 1296
Db      427 AlaGluLysIlePheAsnLysArgGluThrProGluGluArgGluGluArgIleArgArg 446
Qy      1297 GAAGAGAGGAAGGAGGAAGA-----CGT 1323
Db      447 GluThrGluGluLysGluGluArgArgThrGluAspGluGlnLysGluLysGluArg 466
Qy      1324 AATAACGGCAAGAGAAGATTTCAGCTAAGATCTTGGCTGCAGTGGTTGAAGGAAAGC 1383
Db      467 AspArgArgArgHisArgGluMetSerLysLeuLeuAlaThrValValSerGlyGlnLys 486
Qy      1384 AATACGGAAGAGAGAGAGATTTTAGGAAATTAGTTCAGGCCCTAGACAGTCAAGGAAC 1443
Db      487 Gln-----AspArgGlnGlyGlu 493
Qy      1444 CTGGGCAATAGGCCCTCCACACAGGACCAATGTCATATGCTAAGAGAGAGGACAC 1503
Db      494 ---ArgArgArgSerGlnLeuAspArgAspGlnCysAlaTyrCysLysGluLysGlyHis 512
Qy      1504 TGGCAAGAGAACTCCCAAGAAAG-----GGAACAAAGAGCAACCAAG----- 1545
Db      513 TrpAlaLysAspCysProLysLysProArgGlyProArgGlyProArgProGlnThrSer 532
Qy      1546 ATCTAGCTCTAGAAGAA 1563
Db      533 LeuLeuThrLeuAspAsp 538
```

**This Page Blank (uspto)**

GenCore version 5.1.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 16:07:51 ; Search time 39.5271 Seconds  
(without alignments)  
3323.427 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_585\_2156

Perfect score: 2848  
Sequence: 1 ATGGGACAGCGGTGACGAC.....CTCTAGAGAGATTAAGAT 1572

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB.spool/US10723552/runat\_14022006\_125148\_13205/app.query.fasta\_1  
-DB=Published Applications AA Main -QFMT=fascan -SUFFIX=rapbm -MINMARCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-MAXLEN=2000000000 -HOST=abs02p  
-USER=US10723552 @CGN 1.1 602 @runat\_14022006\_125148\_13205 -NCPUS=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREAU=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA Main:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1380.5	48.5	538	4	US-10-677-558-1
2	1379.5	48.4	538	3	US-09-827-822-9
3	967	34.0	957	5	US-10-732-923-13461
4	832.5	29.2	253	4	US-10-670-695-28
5	825	29.0	774	5	US-10-732-923-13451
6	573	20.1	763	5	US-10-732-923-10649
7	553	19.4	405	5	US-10-450-763-38327
8	553	19.4	430	5	US-10-450-763-39297
9	553	19.4	1577	5	US-10-450-763-36834
10	553	19.4	1577	5	US-10-450-763-57482
11	543.5	19.1	611	4	US-10-670-695-32
					Sequence 1, Appli
					Sequence 9, Appli
					Sequence 13461, A
					Sequence 28, Appl
					Sequence 13451, A
					Sequence 10649, A
					Sequence 38327, A
					Sequence 39297, A
					Sequence 36834, A
					Sequence 57482, A
					Sequence 32, Appl

12	542.5	19.0	878	5	US-10-450-763-36843	Sequence 36843, A
13	526.5	18.5	688	5	US-10-450-763-39289	Sequence 39289, A
14	497	17.5	918	5	US-10-732-923-13456	Sequence 13456, A
15	497	17.5	981	5	US-10-732-923-13455	Sequence 13455, A
16	478	16.8	505	5	US-10-450-763-45672	Sequence 45672, A
17	475.5	16.7	208	4	US-10-029-386-33154	Sequence 33154, A
18	439.5	15.4	368	5	US-10-450-763-39293	Sequence 39293, A
19	439.5	15.4	606	5	US-10-450-763-39288	Sequence 39288, A
20	439.5	15.4	606	5	US-10-450-763-39288	Sequence 39288, A
21	439.5	15.4	652	5	US-10-450-763-38335	Sequence 38335, A
22	435.5	15.3	510	5	US-10-450-763-38333	Sequence 38333, A
23	415.5	14.6	202	4	US-10-029-386-32299	Sequence 32299, A
24	400.5	14.1	793	5	US-10-450-763-57559	Sequence 57559, A
25	397.5	14.0	402	5	US-10-450-763-36213	Sequence 36213, A
26	392.5	13.8	997	5	US-10-450-763-39301	Sequence 39301, A
27	392	13.8	1015	5	US-10-450-763-34486	Sequence 34486, A
28	387.5	13.6	1209	5	US-10-450-763-36212	Sequence 36212, A
29	386	13.6	693	5	US-10-450-763-39299	Sequence 39299, A
30	369	13.0	309	5	US-10-450-763-45680	Sequence 45680, A
31	359.5	12.6	245	3	US-09-864-761-36722	Sequence 36722, A
32	355.5	12.5	620	5	US-10-450-763-33954	Sequence 33954, A
33	355.5	12.5	620	5	US-10-450-763-36206	Sequence 36206, A
34	355.5	12.5	620	5	US-10-450-763-36845	Sequence 36845, A
35	355.5	12.5	620	5	US-10-450-763-45682	Sequence 45682, A
36	352.5	12.4	809	5	US-10-450-763-36849	Sequence 36849, A
37	348	12.2	1035	5	US-10-450-763-35678	Sequence 35678, A
38	328	11.5	486	5	US-10-450-763-36846	Sequence 36846, A
39	328	11.5	486	5	US-10-450-763-52359	Sequence 52359, A
40	313.5	11.0	420	4	US-10-637-565-15	Sequence 15, Appl
41	310	10.9	544	5	US-10-450-763-35688	Sequence 35688, A
42	310	10.9	544	5	US-10-450-763-47717	Sequence 47717, A
43	310	10.9	550	5	US-10-450-763-40946	Sequence 40946, A
44	309	10.8	507	5	US-10-450-763-52789	Sequence 52789, A
45	302.5	10.6	390	5	US-10-450-763-39975	Sequence 39975, A

ALIGNMENTS

RESULT 1  
US-10-677-558-1  
; Sequence 1, Application US/10677558  
; Publication No. US2004006972A1  
; GENERAL INFORMATION:  
; APPLICANT: AUDIT, Muriel  
; APPLICANT: COSSET, Francois-Loic  
; TITLE OF INVENTION: CHIMERIC PLASMID COMPRISING A REPLICATIVE RETROVIRAL GENOME AND  
; FILE REFERENCE: 1759.135  
; CURRENT APPLICATION NUMBER: US/10/677,558  
; CURRENT FILING DATE: 2003-10-02  
; PRIOR APPLICATION NUMBER: PCT/FR02/03934  
; PRIOR FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: FR 0114976  
; PRIOR FILING DATE: 2001-11-20  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 538  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: product of gag sequence of pAM plasmid  
US-10-677-558-1

Alignment Scores:  
Pred. No.: 6.8e-95 Length: 538  
Score: 1380.50 Matches: 283  
Percent Similarity: 62.2% Conservative: 69  
Best Local Similarity: 50.0% Mismatches: 141  
Query Match: 48.5% Indels: 73  
DB: 4 Gaps: 13

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-10-677-558-1 (1-538)





QY	1	ATGGGACAGACGGTTCACGACCCCTCTTAGTTTGA	CTCTCGACCATTTGGACTGAAGTTAA	60
Db	1	MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisIleTrpLysAspValGlu	20	
QY	61	TCCAGGGCTCATAAATTTGTCAAGTTCAGGTTAAGAAGGGACCTTGGCAGACTTTCGTGTC	120	
Db	21	ArgLeuAlaHisAsnGlnSerValAspValLysLysArgArgTrpValThrPheCysSer	40	
QY	121	TCTGAATCGCGCAGCATTCGATGTGGATGGCCATCAGAGGGACCTTTAAATTCAGATT	180	
Db	41	AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu	60	
QY	181	ATCCTGGCTGTATAAGCAGTTATTTTCAGACTGGACCGGCTCTCATCCCGATCAGGAG	240	
Db	61	IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal	80	
QY	241	CCCTATATCTTACGTGCGACGAATTTGGCAGAGGATCTCTCCGCCATGGGTTAAACCATCG	300	
Db	81	ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProTrpValLysProPhe	100	
QY	301	CTGAATAAGCCACGAAGCCAGGTCCTCCCGAATTCGGCTCTTGGAGAGAAAACAAACAC	360	
Db	101	ValHis-----ProLysProProProLeuLeuProSer-----	112	
QY	361	TCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTATACCCGAGATTGAGGAGCCACCG	417	
Db	113	-----AlaProSerLeuProLeuGluProProLeuSerThrProProGln	127	
QY	418	-----GTTGGCCGGAAACCCCAA	435	
Db	128	SerSerLeuTyrProAlaLeuThrProSerLeuGlyAlaLysProLysProGlnValLeu	147	
QY	436	-----TCTGTTCCCCACCCCTTATCTGGCA	462	
Db	148	SerAspSerGlyGlyProLeuIleAspLeuLeuThrGluAspProProTyr-----	165	
QY	463	CAGGTCGCGCGAGGGACCTTTGCCCTCCT	495	
Db	166	-----ArgAspProArgProProProSerAspArgAspGlyAspSerGlyGlu	181	
QY	496	GGACTCCGGCGGTGGAGGACCT-----GTCGAGGAGCTCGGAGCCGG	540	
Db	182	AlaThrProAlaGlyGluAlaProAspProSerProMetAlaSerArgLeuArgGlyArg	201	
QY	541	AGSGGCGCCACCCGGAGCGGACAGACAGATTCGGCATTACCGCTCGGACGTACGCGC	600	
Db	202	ArgGluProProValAlaAspSerThrThrSerGlnAlaPheProLeuArgThrGlyGly	221	
QY	601	CCTCCACACCGGGGGGCAATTGCGAGCCCTCCAGTATTGSCCTTTTCTTCTCAGAT	660	
Db	222	-----AsnGlyGln-----LeuGlnTyrTrpProPheSerSerSerAsp	234	
QY	661	CTCTATAATTGGAATAACCATCCCCCTTTCTCGAGGATCCCCAACGCTCACGGGG	720	
Db	235	LeuTyrAsnTrpLysAsnAsnAsnProSerPheSerGluAspProGlyLysLeuThrAla	254	
QY	721	TTGTTGGAGTCCTTATGTTCTCTCACGAGCTACTTCGGATGATGTCAACAGTGCTGT	780	
Db	255	LeuIleGluSerValLeuIleThrHisGlnProThrTrpAspAspCysGlnGlnLeuLeu	274	
QY	781	CAGACACTTTCACAAACGAGGAGCAGAGAGAATTCTATTAGAGCTTAGAAAAATGTT	840	
Db	275	GlyThrLeuLeuThrGlyGluGluLysGlnArgValLeuLeuGluAlaArgLysAlaVal	294	
QY	841	CCTGGGCGCAGCGGCGCACCCACCGGTTGCAAAATGAGATTGATCGGATTTCCCTTTA	900	
Db	295	ArgGlyAspAspGlyArgProThrGlnLeuProAsnGluValAspAlaAlaPheProLeu	314	
QY	901	ACTGCCCGCGGTGGGACTACAAACGCGCTGAAGGTAGGGAGACGCTTGAATCTATCGC	960	
Db	315	GluArgProAspTrpGluTyrThrThrGlnAlaGlyArgAsnHisLeuValHisTyrArg	334	

QY 961 CAGGCTCTGGTGGGGGTCTCCGGGGGCGCTCAAGACGGCCCACTAATTTGGCTCAGGTTA 1020

Db 335 GlnLeuLeuIleAlaGlyLeuGlnAsnAlaGlyArgSerProThrAsnLeuAlaLysVal 354

QY 1021 AGAAGAGTGATGACGAGGACCGAATGAACCCCTCTCTCTTTTCTTGTAGAGGCTCTTGGAA 1080

Db 355 LysGlyIleThrGlnGlyProAsnGluSerProSerAlaPheLeuGluAArgLeuLysGlu 374

QY 1081 GCCTTCAGGCGGTACACCCCTTTTGTATCCCACTTCAGAGGCCCAAAAGAGCCTCAGTGGCT 1140

Db 375 AlaTyrArgArgTyrThrProTyrAspProGluAspProGlyGlnGluThrAsnValSer 394

QY 1141 TTGSCCTTTATAGGACAGTCCAGCTTCGGATATAGAAAGAGCTTCAGACACTGGAGGG 1200

Db 395 MetSerPheIleThrGlnSerAlaProAspIleGlyArgGluArgLeuGluAArg 414

QY 1201 TTCAGAGGCTGAGTTACGTGATCTAGTGAAGGAGGCAGAGAAAGTATATACAAAAGG 1260

Db 415 LeuArgAsnLysThrLeuGlyAspLeuValArgGluAlaGluArgIlePheAsnLysArg 434

QY 1261 GAGCAGAGAAGAAAGGGACAAGAAAGACAGAGAGAAGAGAGGAAAGGAGGAGAAAGA 1320

Db 435 GluThrProGluGluArgGluGluArgGluArgGluArgGluGluArgGluGluArg 454

QY 1321 -----CGTAATAAACCGCAAGAGAAAGATTTC 1347

Db 455 ArgArgThrGluAspGluGlnLysGluLysGluArgAspArgArgHisArgGluMet 474

QY 1348 ACTAAGATCTTGGCTGCAGTGGTGTGAAGGAAAGCAATACGGAAGAGAGAGAGATTTT 1407

Db 475 SerArgLeuLeuAlaThrValValSerGlyGlnArgGlnAspArgGlnGlyGluArg 494

QY 1408 AGGAAATTAGTTCAGGCCCTAGACAGTCAAGGAACTGGGCAATAGGACCCCACTCGAC 1467

Db 495 Arg-----ArgSerGlnLeuAsp 500

QY 1468 AAGGACCAATGTGCATATTGTAAGAAAGAGGACACTGGGCAAGGAACCTGCCCAAG--- 1524

Db 501 CysAspGlnCysThrTyrCysGluGluGlnGlyHisTrpAlaLysAspCysProArgArg 520

QY 1525 ---AAGGAAACAAAGGACCAAG-----ATCCTAGCTCTAGAGAA 1563

Db 521 ProArgGlyProArgGlyProArgProGlnThrSerLeuLeuThrLeuAspAsp 538

RESULT 3

US-10-732-923-13461

; Sequence 13461, Application US/10732923

; Publication No. US20050108791A1

; GENERAL INFORMATION:

; APPLICANT: Edgerton, Michael D

; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES

; FILE REFERENCE: 39-15(52796)C

; CURRENT APPLICATION NUMBER: US/10/732,923

; CURRENT FILING DATE: 2003-12-10

; PRIOR APPLICATION NUMBER: 10/310,154

; PRIOR FILING DATE: 2002-12-04

; NUMBER OF SEQ ID NOS: 24149

; SEQ ID NO 13461

; LENGTH: 957

; TYPE: PRT

; ORGANISM: Gardner-Arnstein feline leukemia oncovirus B

US-10-732-923-13461

Alignment Scores:		
Pred. No.:	8.9e-64	Length:
Score:	967.00	Matches:
Percent Similarity:	49.8%	Conservative:
Best Local Similarity:	37.5%	Mismatches:
Query Match:	34.0%	Indels:
DB:	5	Gaps:

US-10-723-552-3 COPY 585 2156 (1-1572) X US-10-732-923-13461 (1-957)

```
QY 1 ATGGCAGACGGTGACGACCCCTTATGTTTGAATCTCGACCATTCGACCTGAAGTTAAA 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetGlyGlnThrIleThrProLeuSerLeuThrLeuAspHisTrpSerGluValArg 20
QY 61 TCCAGGGCTCATAATTTCTCAGTTTCAAGGACCTTCGACAGCTTCTGTGTC 120
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 AlaArgAlaHisAsnGlnGlyValGluValArgLysLysLysTrpIleThrLeuCysGlu 40
QY 121 TCTGAATGCCGACATTCGATGTTGGATGGCCATCAGAGGGGACCTTTAAATTCGAGATT 180
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41 AlaGluTrpValMetMetAsnValGlyTrpProArgGluGlyThrPheSerLeuAspAsn 60
QY 181 ATCTGGCTGTATAAGCATTTATTTTTCAGACTGGACCCCGCTCTCATCCCGATCAGGAG 240
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 IleSerGlnValGluLysLysIlePheAlaProGlyProTyrGlyHisProAspGlnVal 80
QY 241 CCTATATCTTACGTGGCAAGATTTGGCAGAGATCTCCGCCATCGGTTTAAACCATGG 300
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 ProTyrIleThrThrTrpArgSerLeuAlaThrAspProProSerTrpValArgProPhe 100
QY 301 CTGAATAAGCCCAAGAACCCAGGTCCCGAATTCGTGGCTCTGGAGAGAAAACAACAC 360
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 LeuProProLysProPro-----ThrSer 109
QY 361 TCGGCTGAAAGATCAAGCCCTCTCT-----CATATCTACCCCGAG 402
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 LeuProGlnProLeuSerProGlnProSerAlaProLeuThrSerSerLeuTyrProVal 129
QY 403 ATT-----GAGGAGCCACCGCTTGGCCGGAACCCCAATCTGTTCCC--- 444
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
130 LeuProLysSerAspProProLysProValLeuProProAspProSerSerProLeu 149
QY 445 -----CCACCCCTTATCTGGCACAGGTGCGCCGAGGGGACCC 483
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 IleAspLeuLeuThrGluGluProProProTyr-----ProGlyGlyHisGlyPro 166
QY 484 TTGCCCCCTCTGGAGCTCCG-----CGGTGGAGGACCTCTGTCAGGACTCGGAGC 537
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 -----ProProSerGlyProArgThrProThrAlaSerProIleAlaSerArgLeuArg 184
QY 538 CGGAGGGCGCCACCCCGGAGCGGACAGACGAGATTCGCACATTACCGCTCGCAGCGTAC 597
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 GluArgArgGluAsnPro-----AlaGluGluSerGlnAlaLeuProLeuArg----- 200
QY 598 GGCCCTCCACACCGGGGGGCAATTGACGCCCTCCAGTATTGGCCCTTCTCTCTGCA 657
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 -----GluGlyProAsnAsnArgPro---GlnTyrTrpProPheSerAlaSer 215
QY 658 GATCTCTATAATTGGAATACTAACCATCCCTTCTCGGAGGATCCCAACGCTCACG 717
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 AspleuTyrAsnTrpLysSerHisAsnProProPheSerGlnAspProValAlaLeuThr 235
QY 718 GGGTTGGTGAGTCCCTTATGTTCTCTCACCAGCCCTACTTGGGATGATGTCAACAGCTG 777
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 AsnLeuIleGluSerIleLeuValThrHisGlnProThrTrpAspAspCysGlnGlnLeu 255
QY 778 CTGCAGACACTCTTCCAAACCGAGGAGGAGAGAGAAATCTATTAGAGCTAGAAAAAAT 837
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 LeuGlnAlaLeuLeuThrGlyGluGluArgGlnArgValLeuLeuGluAlaArgLysGln 275
QY 838 GTTCTCGGGCGGAGCGGACCCACGCGTTGCAAAATGAGATTGATCGGATTTCCTCC 897
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
276 ValProGlyGluAspGlyArgProThrGlnLeuProAsnValIleAspGluThrPhePro 295
QY 898 TTAATCTCCCGGTTGGAGCTACAAACGCGTGAAGGTAGGAGAGCTTGAAATCTAT 957
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 LeuThrArgProAsnTrpAspPheAlaThrProAlaGlyArgGluHisLeuArgLeuTyr 315
QY 958 GCCAGGCTCTGGTGGCGGTCTCGGGCGGCTCAAGACGCCCACTAAATTTGGCTAAG 1017
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
316 ArgGlnLeuLeuAlaGlyLeuArgGlyAlaAlaArgProThrAsnLeuAlaGln 335
QY 1018 GTAAGAGAAGTGTATGAGGAGCCGAATGAACCCCTCTGTTTCTTTGAG----- 1068
```

```
Db 336 ValLysGlnValValGlnGlnLysGluGluThrProAlaAlaAlaAlaAspGlyThr 355
QY 1068 -----
Db 356 MetGlyPheSerSerGluLeuCysSerProGlnGlyHisGlyAlaGluGlnGlnMetGln 375
QY 1069 -----AGGCTCTTGGAGCCCTTCAGGCGGTACACC----- 1098
Db 376 GluAlaGluLeuArgLeuGluGlyMetArgLysTrpMetAlaGlnArgValLysSer 395
QY 1098 -----
Db 396 AspArgGluTyrAlaGlyLeuLeuHisHisMetSerLeuGlnAspGlyGlyArgGly 415
QY 1099 -----CCTTTGATCCACCTTCAGAGCGCCAAAGCCCTCAGTGGCTTTGGCCTTTATA 1152
Db 416 ThrGlyProTyrSerProIleSerGlnSerTrp---AlaGluIleThrSerGlnThrGlu 434
QY 1153 GGACAGTCAGCTTGGATATTAGAAAGAAGCTTCAGAGACTTGAAGGTTACAGGAGCT 1212
Db 435 GlyLeuSerArgLeu---LeuArgGlnHisAlaGluAspLeuAsnSerGlyProLeuSer 453
QY 1213 GAGTTAGCTGATCTAGTCAAGGCGCAGAGAAAGTATATTACAAAAGGAGACACAAAGAA 1272
Db 454 LysLeuGlyLeuLeuIleArgGluArgGlnGlnLeu-----ArgLysThrTyrSer 470
QY 1273 GAAGGGAACAAGAAAGAGAGAGAA----- 1299
Db 471 GluGlnTrpGlnGlnLeuGlnGlnGluLeuThrLysThrHisAsnGlnAspIleGluLys 490
QY 1300 -----AGAGAGAAAGGAGGAGAAAGACGCTAATAAACGG 1332
Db 491 LeuLysSerGlnTyrArgAlaLeuAlaArgAspSerAlaGlnAlaArgArg---LysTyr 509
QY 1333 CAAGAGAAAGAAATGACTAAG----- 1353
Db 510 GlnGluAlaSerLysAspLysAspArgAspLysAlaLysLeuGluGlnLeuGlyProGly 529
QY 1354 -----ATCTGGCTGCAGTGTTCAGAGGAAAGCAATACCGAAAGA 1395
Db 530 GluProProValLeuLeuLeuGlnAspAspArgHisSerThrSerSerSerGluGln 549
QY 1396 GAGAGAGATTTAGGAAAATTTAGTCTAGGCCCTAGACAGTCAAGGAAACCTGGGCAATAGG 1455
Db 550 GluArgGlu----- 555
QY 1456 ACCCA 1461
Db 556 ThrPro 557
RESULT 4
US-10-670-695-28
; Sequence 28, Application US/10670695
; Publication No. US20040058316A1
; GENERAL INFORMATION:
; APPLICANT: Jensen, Wayne A.
; APPLICANT: Lappin, Michael R.
; APPLICANT: Rosen, David K.
; APPLICANT: Andrews, Janet S.
; TITLE OF INVENTION: USE OF RECOMBINANT ANTIGENS TO DETERMINE THE IMMUNE
; TITLE OF INVENTION: STATUS OF AN ANIMAL
; FILE REFERENCE: DI-9-1
; CURRENT APPLICATION NUMBER: US/10/670,695
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: 09/521,738
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Feline leukemia virus
```

## US-10-670-695-28

## Alignment Scores:

Pred. No.: 8,36e-54 Length: 253  
Score: 832.50 Matches: 159  
Percent Similarity: 77.6% Conservative: 38  
Best Local Similarity: 62.6% Mismatches: 50  
Query Match: 29.2% Indels: 7  
DB: 4 Gaps: 2

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-10-670-695-28 (1-253)

QY 580 TTACGGCTGCGACAGTACGGCCCTCCACACGGGGGCCAAATGACGCCCTCCAGTAT 639  
DB 1 MetProLeuArg-----GluGlyProAsnAsnArgPro---GlnTyr 13  
QY 640 TGGCCCTTTCTTCGACAGTCTCTATAATGGAATACTAACCATCCCTCTCCGAG 699  
DB 14 TrpProPheSerAlaSerAspLeuTyrAsnTrpLysSerHisAsnProProPheSerGln 33  
QY 700 GATCCCCAACCCCTCACGGGGTGGTGGAGTCCCTTATGTCTCTCACAGCCTACTTGG 759  
DB 34 AspProValAlaLeuThrAsnLeuIleGluSerIleLeuValThrHisGlnProThrTrp 53  
QY 760 GATGATGTTCAACAGCTCTCCAGACACTCTTCAACACCGGAGGAGAGAGAAATCTA 819  
DB 54 AspAspCysGlnGlnLeuLeuGlnAlaLeuLeuThrGlyGluGluArgGlnArgValLeu 73  
QY 820 TTACAGGCTAGAAAAATGTTCTTGGGCGCCGACGGGACCCACGCGTTCGAAATGAG 879  
DB 74 LeuGluAlaArgLysGlnValProGlyGluAspGlyArgProThrGlnLeuProAsnVal 93  
QY 880 ATTGACATGGGATTTCCTTAACTCGCCCGGTTGGGACTCAACACGGCTGAAGTAGG 939  
DB 94 IleAspGluAlaPheProLeuThrArgProAsnTrpAspPheAlaThrProAlaGlyArg 113  
QY 940 GAGAGCTTGAAATCTATCCGACGCTCTGTGGCGGTCTCCGGGGCGCTCAAGACGG 999  
DB 114 GluHisLeuArgLeuTyrArgGlnLeuLeuAlaGlyLeuArgGlyAlaAlaArgArg 133  
QY 1000 CCCACTAATTTGGCTTAAGTAAGAGTAGTGACGAGGACCGAATGACCCCTCTGTT 1059  
DB 134 ProThrAsnLeuAlaGlnValLysGlnValGlnGlyLysGluGluThrProAlaSer 153  
QY 1060 TTTCTTGAGAGCTCTTGGAGCCCTCAGGCGGTACACCCCTTTGATCCACCTCAGAG 1119  
DB 154 PheLeuGluArgLeuLysGluAlaTyrArgMetTyrThrProTyrAspProGluAspPro 173  
QY 1120 GCCCAAAAGCTCAGTGGCTTTGGCCCTTTATAGGACAGTCAGCCTTGGATATPAGAAAG 1179  
DB 174 GlyGlnAlaAlaSerValIleLeuSerPheIleTyrGlnSerSerProAspIleArgAsn 193  
QY 1180 AAGCTTCAGAGCTGGAAGGTTACAGAGGCTGAGTACGTGATCTAGTGAAGAGGCA 1239  
DB 194 LysLeuGlnArgLeuGluGlyLeuGlnGlyPheThrLeuSerAspLeuLeuLysGluAla 213  
QY 1240 GAGAAAGTATATTACAAAGGAGACAGAGAGGAAAGGAAACAAAGAGAGAGAA 1299  
DB 214 GluLysIleTyrAsnLysArgGluThrProGluGluArgGluGluArgLeuTyrGlnArg 233  
QY 1300 AGAGAGAAAGGAGAGAAAGCTGTAATAAAGCGGCAAGAGAG 1341  
DB 234 GlnGluGluArgAspLysArgHisLysGluMetThrLys 247

## RESULT 5

US-10-732-923-13451  
; Sequence 13451, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923

; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 13451  
; LENGTH: 774  
; TYPE: PRT  
; ORGANISM: Feline sarcoma virus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(774)  
; OTHER INFORMATION: unsure at all Xaa locations  
; US-10-732-923-13451

Alignment Scores:  
Pred. No.: 3.93e-53 Length: 774  
Score: 825.00 Matches: 191  
Percent Similarity: 53.1% Conservative: 50  
Best Local Similarity: 42.1% Mismatches: 135  
Query Match: 29.0% Indels: 78  
DB: 5 Gaps: 14

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-10-732-923-13451 (1-774)

QY 1 ATGGGACAGCGGTGACGACCCCTTTAGTTGACTCTCGACCAATTGGACTGAAGTTAAA 60  
DB 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuAspHisTrpSerGluValArg 20  
QY 61 TCCAGGCTCATATTTGTCAAGTTAGTAAAGAGGACCTTGGCAGACTTCTGTGTC 120  
DB 21 AlaArgAlaHisAsnGlnGlyValGluValArgLysLysIleThrLeuCysLys 40  
QY 121 TCTGAATGGCGCACATTCGATGTTGGATGGCCATCAGAGGGACCTTTAATTCGAGATT 180  
DB 41 AlaGluTrpValMetMetAsnValGlyTrpProArgGluGlyThrPheSerLeuAspAsn 60  
QY 181 ATCTGCTGCTTAAAGCAGTTATTTTCAGACTGGACCCGCTCTCATCCGATCAGAG 240  
DB 61 IleSerGlnValLysLysIlePheAlaProGlyProHisGlyHisProAspGlnVal 80  
QY 241 CCCTATATCTTACGTGGCAAGATTGGCAGAGGATCTCGCCATCGGTAAACATGG 300  
DB 81 ProTyrIleThrThrTrpArgSerLeuAlaThrAspProSerTrpValArgProPhe 100  
QY 301 CTGAATAAGCAAGAACCCAGGTCCCGGAATTCGTGCTCTTGAGAGAGAAAAACAACAC 360  
DB 101 LeuProProLysProProThrPro----- 109  
QY 361 TCGGCTGAAAAGTCAAGCCCTCTCCT-----CATATCTACCCGAG 402  
DB 110 LeuProGlnProLeuSerProGlnProSerAlaProLeuThrSerSerLeuTyrProVal 129  
QY 403 ATT-----GAGGAGCCACCGCTTGGCCGGAACCCCAATCTGTTCCC--- 444  
DB 130 ValProLysProAspProProLysProProValLeuProProAspProSerSerProLeu 149  
QY 445 -----CACACCCCTTATCTGSCACAGGTGCGCGAGGGAGCCC 483  
DB 150 IleAspLeuLeuThrGluGluProProProTyr-----ProGlyGlyHisGlyProPro 167  
QY 484 TTTGCCCCCTCTCGAGCTCCGGCGGTGGAGGACCTGCTGCAGGAGCTCGAGCGCGAGG 543  
DB 168 ProSerGlyProArgThrProAlaAlaSerProIleValSerArgLeuArgGluArgArg 187  
QY 544 GGGCCACCCCGGAGGAGCAGACAGATTCGACATTACCGTTCGCGCAGTACGGCCCT 603  
DB 188 -----GluAsnProAlaGluGluSerGlnAlaLeuProLeuArg----- 200  
QY 604 CCCACACCGGGGGCCAAATTGACGCCCTCCAGTATTGGCCCTTTCTTCTTCGAGATCTC 663  
DB 201 -----GluGlyProAsnAsnArgPro---GlnTyrTrpProPheSerAlaSerAspLeu 217  
QY 664 TATAATTGGAAAACTAAACCATCCCTTTCTCGGAGGATCCCAACAGCCTTCACGGGGTTG 723



```
Db 306 ArgProArgTyrPheLeuLeuLysAsnAspGlyThrPheIleGlyTyrLysGluArgPro 325
Qy 1003 ACTAATTGGCTAAGTAAGA-----GAGGTG 1029
Db 326 GlnAspValAspGlnArgGluSerProLeuAsnAsnPheSerValAlaGlnCysGlnLeu 345
Qy 1030 ATGACGGGACCGAATGAACCCCTCTCTTTTCTTGAGAGGCTCTGGAA----- 1080
Db 346 MetLysThrGluArgProArgProAsnThrPheIleIleArgCysLeuGlnTrpThr 365
Qy 1081 -----GCCTTCAGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAA 1128
Db 366 ValIleGluArgThrPheHisValGluThrPro-----GluGluArgGluGlu 381
Qy 1129 GCCTCAGTGGCTTTGGCCCTTTATAGACAGTCAGCTTGGATATTAGAAAGAGCTTCAG 1188
Db 382 TrpAlaThrAlaIleGlnThrValAla----- 390
Qy 1189 AGACTGGAGGTTTACAGGCGCTGAGTTACGTGATCTAGTGAAGGAGCGCAGAAAAGTA 1248
Db 391 -----AspGlyLeu----- 393
Qy 1249 TATTACAAAAGGGAGACAGAGAAGAAAGGGAACAAAGAAAA-----GAG 1293
Db 394 -----LysArgGlnGluGluThrMetAspPheArgSerGlySerProSerAspAsn 411
Qy 1294 AGAAAGAGAGGAAGGAGGAAGAGAGCTAATAACCGCAAGAGAGAAGAAATTTGACTAAG 1353
Db 412 SerGlyAlaGluGluMetGlu-ValSerLeuAlaLysProLysHisArgValThrMetAs 431
Qy 1354 ATCTTGCTGCTGAGTGGTTGAAGGGAAGAAACAAATACGGAAGAGAGAGAGATTTAGAAA 1413
Db 431 nGluPhe---GluTyrLeuLys-----LeuLeuGlyLy 441
Qy 1414 ATTAGTCCAGGCCCTAGACAGTCAGGAGCACTGGGCAATAGGACCCCTCAGCAAGAC 1473
Db 441 s-----GlyThrPheGly----- 445
Qy 1474 CAATGTGCATATTGAAGAAAGAGACACTGGGCAA 1510
Db 446 -LysValIleLeuValLysGluLysAlaThrGlyArg 457

RESULT 7
US-10-450-763-38327
; Sequence 38327, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 38327
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (106)..(311)
; OTHER INFORMATION: Gag p30 core shell protein domain identified by PFam,
; OTHER INFORMATION: accession name Gag_p30, E-value=1.2e-67, PFam score of 238.2
US-10-450-763-38327
Alignment Scores:
Pred. No.: 9.18e-33 Length: 405
```

```
Score: 553.00 Matches: 151
Percent Similarity: 48.6% Conservative: 64
Best Local Similarity: 34.2% Mismatches: 149
Query Match: 19.4% Indels: 78
DB: 5 Gaps: 19

US-10-723-552-3_copy_585_2156 (1-1572) x US-10-450-763-38327 (1-405)
Qy 298 TGGCTGAATTAAGCCCAAGAAAGCCAGGTCCTCCGAATTCCTGGCTCTTGGAGAGAAAAACAAA 357
Db 14 TrpLeuAsnThrGlyLysGlu-----LeuAlaLeuGlyValTrpThrPhe 28
Qy 358 CACTCGCGCTGAAAAAGTCAAGCCC-----TCTCCT-----CATATCTACCCCGAG 402
Db 29 GluTrpGluArgGlySerProLeuLeuSerProArgGlyLeuHisLeuHis----- 46
Qy 403 ATTGAGGAGCCACCGGCTTGGCGGAACCCCAATCTGTCTCCCCACCCCTTATCTGCGCA 462
Db 47 -----LysThrAsnIleProArgProProArgValGly 57
Qy 463 CAG-----GCTGCCGCGAGGGGACCCTTTGCCCTCTCTGGAGCTCCGCGGTGGAGGA 516
Db 58 LysArgGlyGlyGluAlaSerGlyGlu---ThrProPro----- 69
Qy 517 CCTGCTCAGGACTCGAGCCGAGGGGCGCCACCCCGAGCGGAGCAGACGAGATCGCG 576
Db 70 LeuAlaAlaGlyLeuArgProLysThrGly-----Ile 80
Qy 577 ACATTACCGCTGGCAGC-----TACGGCCCTCCACACCGGGGGGCCAATTG----- 624
Db 81 GlnMetProLeuArgGluGlnTrpTyrThrGlyIleAspGluAspGlyHisMetValGlu 100
Qy 625 ---CAGCCCTCCAGTATTGGCCCTTTCTTCTGCAGATCTCTATAATTTGAAAACTAAC 681
Db 101 ArgArgValPheValTyrGlnProPhePheSerAlaAspLeuLeuAsnTrpLysAsnAsn 120
Qy 682 CATCCCTCTTCTCGGAGGATCCCAACGCTCAACGGGTTGGTGGAGTCCCTTATGTC 741
Db 121 ThrProSerTyrThrGluLysProGlnAlaLeuIleAspLeuLeuGlnThrIleIleGln 140
Qy 742 TCTCACCAGCTACTTGGGATGATTGTCAACAGCTGCTGCAGACACCTTTCACACCCGAG 801
Db 141 ThrHisAsnProThrTrpAlaAspCysHisGlnLeuLeuMetPheLeuPheAsnThrAsp 160
Qy 802 GAGCGAGAGAGAAATCTATTAGAGGCTAGAAAA-----AATGTTCTTGGGCGC 849
Db 161 GluArgArgValLeuGlnAlaAlaThrLysTrpLeuGluGluHisAlaProAlaAsp 180
Qy 850 GACGGCGACCCACCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTAACTCGCCCC 909
Db 181 TyrGlnAsnPro-----GlnGluTyr-ValArgThrGlnLeuProGlyThrAspPro 197
Qy 910 GGTGGGACTACAACACCGCTGAAGGTAGGAGAGCTTGAATAATCTATCGCAGGCTCTG 969
Db 198 GlnTrpAspProAsnGluArgGluAspMetGlnArgLeuAsnArgTyrArgGluAlaLeu 217
Qy 970 GTGGCGGCTCTCGGGCGCGCTCAACAGCGCCCACTAATTTGGCTAAGTAAGAGAGTG 1029
Db 218 LeuGluGlyLeuLysArgGlyAlaGlnLysAlaThrAsnValAla----- 232
Qy 1030 ATGACGGAGCCGAATGAACCCCTCTGTTTTTTTGTGAGAGGCTCTTGGAGGCTTCAGG 1089
Db 233 IleGlnGlyLysGluGluSerProAlaGlnPheTyrGluArgLeuCysGluAlaTyrCys 252
Qy 1090 CGGTACACCCCTTTGATCCCACTCAGAGGCCCAAAAGCCCTCAGTGGCTTGGGCTTTT 1149
Db 253 MetTyrThrProPheAspProAsnSerProGluAsnGlnArgMetIleAsnMetAlaLeu 272
Qy 1150 ATAGGACAGTCACCTTGGATATTAGAAAGAGCTTTCAGAGACTGGNAGGTTTACAGAG 1209
Db 273 ValSerGlnSerAlaGluAspIleArgArgLeuGlnLysGlnAlaGlyPheAlaGly 292
Qy 1210 GCTGAGTTACGTGATCTAGTGAAGGAGCGCAGAGAAAGTATATTACAAAAGGAGACAGAA 1269
```







Db 426 AsnLysCysProGlnLeuLysGlyLysGlnGlyAsp-----SerGluGlnGluAsp 442  
QY 1567 AAAGAT 1572  
Db 443 ProAsp 444

RESULT 10  
US-10-450-763-57482  
; Sequence 57482, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 57482  
; LENGTH: 1577  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (472)..(488)  
; OTHER INFORMATION: Eukaryotic and viral aspartyl proteases proteins domain  
; OTHER INFORMATION: identified by eMATRIX, accession number BL00141A, p-value=4.150e-  
; OTHER INFORMATION: 10, raw score of 12.10  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (156)..(361)  
; OTHER INFORMATION: Gag p30 core shell protein domain identified by PFam,  
; OTHER INFORMATION: accession name Gag\_p30, E-value=1.2e-67, Pfam score of 238.2  
US-10-450-763-57482

Alignment Scores:  
Pred. No.: 1,25e-32 Length: 1577  
Score: 553.00 Matches: 151  
Percent Similarity: 48.6% Conservative: 64  
Best Local Similarity: 34.2% Mismatches: 149  
Query Match: 19.4% Indels: 78  
DB: 5 Gaps: 19

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-10-450-763-57482 (1-1577)  
QY 298 TGGCTGAATAAGCCAAAGAACCCAGGTCCTCCCGAATTCGTGCTCTTGGAGAGAAAAACAAA 357  
Db 64 TrpLeuAsnThrGlyLysGlu-----LeuAlaLeuGlyValTrpThrPhe 78  
QY 358 CACTCGGCTGAATAAGTCAAGCC-----TCTCCT-----CATATCTACCCCGAG 402  
Db 79 GluThrTrpGluArgGlySerProLeuLeuSerProArgGlyLeuHisLeuHis----- 96  
QY 403 ATTGAGGAGCCACCGGCTTGGCGGGAACCCCAATCTGTTCCCAACCCCTTATCTGGCA 462  
Db 97 -----LysThrAsnIleProArgProProArgValGly 107  
QY 463 CAG-----GFTGCCGCGAGGGGACCTTTTGCCCTCTCTGGAGCTCCCGCGGTGGAGGA 516  
Db 108 LysArgGlyGlyGluAlaSerGlyGlu---ThrProPro----- 119  
QY 517 CCTGCTGAGGACTCGGAGCGGAGGGGCGCCACCCCGGAGCGGACAGACGAGATCGCG 576  
Db 120 LeuAlaAlaGlyLeuArgProLysThrGly-----Ile 130  
QY 577 ACATTACCGCTGCGCACG-----TACGGCCCTCCCAACACACCGGGGGGCCAATTG----- 624

Db 131 GlnMetProLeuArgGluGlnTrpTyThrGlyIleAspGluAspGlyHisMetValGlu 150  
QY 625 ---CAGCCCTCCAGTATTGGCCCTTTTCTCTCGAGATCTCTATAATTGAAACTAAC 681  
Db 151 ArgArgValPheValTyThrGlnProPhePheSerAlaAspLeuLeuAsnTrpLysAsnAsn 170  
QY 682 CATCCCTTTCTCGGAGGATCCCAACGCCCTCACGGGGTGGTGGAGTCCCTTATGTTTC 741  
Db 171 ThrProSerTyThrGluLysProGlnAlaLeuIleAspLeuLeuGlnIleIleGln 190  
QY 742 TCTCACCAGCCTACTTGGGATTTGTCAACAGCTGCTGCAGACACTCTTCACAAACCGAG 801  
Db 191 ThrHisAsnProThrTrpAlaAspCysHisGlnLeuLeuMetPheLeuPheAsnThrAsp 210  
QY 802 GAGCAGAGAGAAATCTATTAGAGCTAGAAAA-----AATGTTCTCGGGGCC 849  
Db 211 GluArgArgValLeuGlnAlaAlaThrLysTrpLeuGluGluHisAlaProAlaAsp 230  
QY 850 GACGGCGACCCACCGCGTTGCAAAATGAGATTGACATGGGATTTCCCTTAACTCGCCCC 909  
Db 231 TyrGlnAsnPro-----GlnGluTyValArgThrGlnLeuProGlyThrAspPro 247  
QY 910 GGTGGGACTACAACACCGCTGAAGTAGGAGAGCTTTGAAAAATCTATCGCCAGCTCG 969  
Db 248 GlnTrpAspProAsnGluArgGluAspMetGlnArgLeuAsnArgTyArgGluAlaLeu 267  
QY 970 GTGGGGGTCTCCGGGGCCCTCAAGACGGCCCACTAATTGGTGTAGGTAAGTAAGAGATG 1029  
Db 268 LeuGluGlyLeuLysArgGlyAlaGlnLysAlaThrAsnValAla----- 282  
QY 1030 ATGCAGGACCGAATGAACCCCTCTGTTTTCTTGAGAGGCTCTTGAAGCCCTTCAGG 1089  
Db 283 IleGlnGlyLysGluGluSerProAlaGlnPheTyThrGluArgLeuCysGluAlaTyCys 302  
QY 1090 CGGTACACCCCTTTTGATCCCACTCCAGAGCGCCCAAAAGCCTCAGTGGCTTTGGCCCTT 1149  
Db 303 MetTyThrProPheAspProAsnSerProGluAsnGlnArgMetIleAsnMetAlaLeu 322  
QY 1150 ATAGACAGTCAAGCTTGGATATTAGAAAGCTTCAGACACTGGAAGGGTTACAGAG 1209  
Db 323 ValSerGlnSerAlaGluAspIleArgArgLysLeuGlnLysGlnAlaGlyPheAlaGly 342  
QY 1210 GCTGAGTTACGTGATCTAGTGAAGGAGGAGAGAGAAAGTATATTACAAAGGAGACAGAA 1269  
Db 343 MetAsnThrSerGlnLeuLeuGluIleAlaAsnGlnValPheValAsnArgAspAlaVal 362  
QY 1270 GAAGAAAGGGAAACAAAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1329  
Db 363 SerProLysGluAsnArgArgGlu-----AsnGluHisGlnValArgArgAsn--- 378  
QY 1330 CGGCAAGAGAAGAAATTTGACTAAGATCTTGGCTGCGAGTGTGAAGGAAAGCAATACG 1389  
Db 379 -----AlaAspLeuLeuAlaAlaIleThrGlyValProProLys 392  
QY 1390 GAAAGAGAGAGAGATTTTAGGAAAATTAAGTTCAGCCCTAGACAGCTCAGGAAACCTGGGC 1449  
Db 393 -----ArgGlnGlyLysGlyProGlyLysGluIleGlnProGly 406  
QY 1450 AATAGGACCCCACTCGACAAGGACCAATGTGCATATTGTAAGAAAGAGGACACTGGGCA 1509  
Db 407 CysGlnSer---LeuGlnArgAsnGlnCysAlaTyCysLysGluIleGlyHisTrpLys 425  
QY 1510 AGGAACTGCCCAAG---AAGGAAACAAAGGACCAAGGATCTAGTCTTAGAAGAGAT 1566  
Db 426 AsnLysCysProGlnLeuLysGlyLysGlnGlyAsp-----SerGluGlnGluAsp 442  
QY 1567 AAAGAT 1572  
Db 443 ProAsp 444

RESULT 11  
US-10-670-695-32

```

: Sequence 32, Application US/10670695
: Publication No. US20040058316A1
: GENERAL INFORMATION:
: APPLICANT: Jensen, Wayne A.
: APPLICANT: Lappin, Michael R.
: APPLICANT: Rosen, David K.
: APPLICANT: Andrews, Janet S.
: TITLE OF INVENTION: USE OF RECOMBINANT
: TITLE OF INVENTION: STATUS OF AN
: FILE REFERENCE: DI-9-1
: CURRENT APPLICATION NUMBER: US/10670695
: CURRENT FILING DATE: 2003-09-25
: PRIOR APPLICATION NUMBER: 09/521,733
: PRIOR FILING DATE: 2000-03-09
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 32
: LENGTH: 611
: TYPE: PR1
: ORGANISM: Feline leukemia virus
: US-10-670-695-32

```

Alignment Scores:	5.21e-32	Length:	611
Pred. No.:	543.50	Matches:	114
Score:	543.50	Conservative:	30
Percent Similarity:	67.0%	Mismatches:	46
Best Local Similarity:	53.0%	Indels:	25
Query Match:	19.1%	Gaps:	4
DB:	4		

US-10-723-552-3 COPY 585 2156 (1-1572) x US-10-670-695-32 (1-611)

940	QY	GAGAGCTTGA	AAATCTATCGC	AGGCTCTGTG	TGGGGGTCTCC	GGGGGGCCCTCAAG	CGG 999	
2	Db	GLHISleu	ArgLeu	TrpArg	GlnLeu	LeuLeuAlaGlyLeu	ArgGlyAlaLeuArgHis 21	
1000	QY	CCCACTAA	TTTGGCTAAG	GAAGAAGTGA	TGCAGG	GACCGAATGAAC	CCCCCTCTGTT 1059	
22	Db	ProThr	AsnLeu	AlaGln	ValLysGln	Phel	LeuGlnGlyLysGluThrPro	AlaSer 41
1060	QY	TTTCTTG	AGAGGCTCTT	TGGAAGCTT	CAGCGGT	TACACCCCTTT	TGATCCCACCTCAGAG 1119	
42	Db	PheLeu	GluArgLeu	LysGlu	AlaTrpArg	MetTyrThrPro	TrpArgProGluAspPro 61	
1120	QY	GCCCAAAA	AGCCCTCAG	TGGGCTTT	TATAGGACAG	TGCAGCTTGG	TATAGAAAG 1179	
62	Db	GlyGln	AlaAlaSer	ValIleLeu	SerPhe	IleTyrGln	SerSerProAspIleArg	Asn 81
1180	QY	AACTTTC	CAGAGACT	TGGAAGG	TTTACAGG	AGGCTGAGTTAC	TGATCTAGTGAAGGAGCA 1239	
82	Db	LysLeu	GlnArgLeu	GluGlyLeu	GlnGlyPhe	ThrLeu	SerAspLeuLeuLysGlu	Ala 101
1240	QY	GAGAAAG	TATATTACA	AAAGGG	ACACAGAGA	GAAGAAGGGAAC	AAAGAAAGACAGAGAA 1299	
102	Db	GLuLys	IleTyrAsn	LysArgGlu	ThrProGlu	GluArgGlu	GluArg----- 117	
1300	QY	AGAGAGA	AAAGGAGG	AAAGACCT	TAAATAC	CGGCAAGAGAG	AAGATTTGCTAAGATCTTG 1359	
118	Db	LeuTrp	GlnArgGln	GluGluArg	AspLysArg	HisArgHis	LysGluMetThrLys	ValLeu 137
1360	QY	GCTGCAG	TGTTGAA	GGGAAAAG	CAATATAC	CGAAGAGAGAGA	-----GATTTT 1407	
138	Db	AlaThr	ValVal	AlaGln	AsnArgAsp	LysAspArgGlu	GluSerLysLeuGlyAsp	Gln 157
1408	QY	AGAAAA	ATTAGGTC	CAGGCCCT	TAGACACT	CAGGGAACCT	TGGGCAATAGGACCCCACTCGAC 1467	
158	Db	ArgLys	Ile-----	-----	-----	-----	-----ProLeuGly 163	
1468	QY	AAGGAC	CAATGTGC	ATATTCT	TAAAGA	AGAGGACACT	GGGCAAGGAAGTCCCCCAAGAG 1527	
164	Db	LysAsp	GlnCys	AlaTrpCys	LysGluLysGly	HisTrpVal	ArgAspCysProAsnArg 183	
1528	QY	GGAAAC	AAAGGACCA	-----	-----	AGGATCCT	AGCTCTAGAAGAA 1563	

```

Db      184  ProArgLysLysProAlaAsnSerThrLeuLeuAsnLeuGluAsp 198
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 12
US-10-450-763-36843
; Sequence 36843, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/549,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 36843
; LENGTH: 878

```

```

/ ORGANISM: Homo sapiens
/ TYPE: PRT
/
/ FEATURE:
/
/ NAME/KEY: DOMAIN
/ LOCATION: (391)..(407)
/ OTHER INFORMATION: Eukaryotic and viral a
/ OTHER INFORMATION: identified by eMATRIX,
/ OTHER INFORMATION: 10, raw score of 12.10
/
/ FEATURE:
/
/ NAME/KEY: DOMAIN
/ LOCATION: (75)..(280)
/ OTHER INFORMATION: Gag P30 core shell prot
/ OTHER INFORMATION: accession name Gag_p30
/
/ US-10-450-763-36843

```

Alignment Scores:		
Pred. No.:	6.72e-32	Length:
Score:	542.50	Matches:
Percent Similarity:	50.4%	Conservative:
Best Local Similarity:	35.5%	Mismatches:
Query Match:	19.0%	Indels:
DB:	5	Gaps:
		878
		139
		58
		137
		57
		15

US-10-723-552-3 COPY 585 2156 (1-1572) x US-10-450-763-36843 (1-878)

436	QY	TCGTGTTCCCCACCCCTTATCTGGCACAG-----CGTGCCGCGAGGGGACCCCTTTGCC	489
18	Db	AsnIleProArgProProArgValGlyLybArgGlyGlyGluAlaSerGlyGlu---	36
490	QY	CCTCTGGAGACTCCGGCGGTGGAGGACCTCTGTCAGGAGACTCGAGCCAGGGGGGCC	549
37	Db	ProPro-----LeuAlaAlaGlyLeuArgProLybThrGly---	48
550	QY	ACCCCGGCGGACAGACGAGATCGCGACATTACCGCTGCGCACG-----TAGGCCCT	603
49	Db	-----IleGlnMetProLeuArgGluGlnTrpIleThrGly	60
604	QY	CCCACACCGGGGGGCAATTG-----CAGCCCTCCAGTATGGCCCTTTTCTTCT	654
61	Db	IleAspGluaspGlyHisMetValGluArgValPheValTyrGlnProPhePheSer	80
655	QY	GCAGATCTCTATTAATTGGAAAACTACCAATCCCCCTTTCTCGGAGGATCCCCACGCCTC	714
81	Db	AlaAspLeuLeuAsnTrpLysAsnAsnThrProSerTyrThrGluLysProGlnAlaLeu	100
715	QY	ACGGGGTTGGTGAGTCCCTTATGTCTCTCACCGAGCTACTTGGGATGATTGTCAACAG	774
101	Db	IleAspLeuGlnThrIleIleGlnThrHisAsnProThrTrpAlaAspCysHisGln	120
775	QY	CTGCTCGACACACTCTTTCACAAACCGAGGCGGAGAGAGAAATTCATTAGAGGCTAGAAAA	834

```
Db 121 LeuLeuMetPheLeuPheAsnThrAspGluArgArgValLeuGlnAlaAlaThrLys 140
Qy 835 -----AATGTCCTGGGCGCGACGGGCGACCCACCGGTTGCAAAATGAGATT 882
Db 141 TrpLeuGluGluHisAlaProAlaAspTyrGlnAsnPro-----GlnGluTyrVal 157
Qy 883 GACATGGGATTTCCCTTAATCGCCCGGTTGGGACTACACACGGCTGAAGGTAGGAG 942
Db 158 ArgThrGlnLeuProGlyThrAspProGlnTrpAspProAsnGluArgGluAspMetGln 177
Qy 943 AGCTTGAATAATCATCGCAGGCTCTGGTGGCGGCTCTCCGGCGCCTCAAGACGGCCC 1002
Db 178 ArgLeuAsnArgTyrArgGluAlaLeuLeuGluGlyLeuLysArgGlyAlaGlnLysAla 197
Qy 1003 ACTAATTTGGCTAAGTAAGAAAGTGAAGGACCGCAATGAACCCCTCTCTTTT 1062
Db 198 ThrAsnValAla-----IleGlnGlyLysGluGluSerProAlaGlnPhe 212
Qy 1063 CTTGAGAGGCTTGGAAAGCCTTACAGCGGTACACCCCTTTGTATCCACCTCAGAGGCC 1122
Db 213 TyrGluArgLeuCysGluAlaTyrCysMetTyrThrProPheAspProAsnSerProGlu 232
Qy 1123 CAATAAGCCTCAGTGGCTTTGGCCTTTATAGGACAGTCAGCCTTGATATTAGAAAGAG 1182
Db 233 AsnGlnArgMetIleAsnMetAlaLeuValSerGlnSerAlaGluAspIleArgArgLys 252
Qy 1183 CTTGAGAGCTGGAAGGTTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAGCGCAGAG 1242
Db 253 LeuGlnLysGlnAlaGlyPheAlaGlyMetAsnThrSerGlnLeuLeuGluIleAlaAsn 272
Qy 1243 AAAGTATTATACAAAGGAGACAGAAGAAAGGACAAAGAAAGAAAGAGAGAGAGAAGA 1302
Db 273 GlnValPheValAsnArgAspAlaValSerProLysGluAsnArgArgGlu----- 289
Qy 1303 GAGGAAAGGAGAAAGAGCTAATAAACGGCAAGAGAGAATTTGACTAAGATCTTGGCT 1362
Db 290 AsnGluHisGlnValArgArgAsn-----AlaAspLeuLeuAla 302
Qy 1363 CGAGTGGTTGAAGGAAAGCAATACGGAAGAGAGAGAGATTTTAGGAAATTAGGTCA 1422
Db 303 AlaAlaIleThrGlyValProProLys-----ArgGlnGlyLysGly 316
Qy 1423 GGCCTTAGACGTACGGGNACTGGGCAATAGGACCCACACGACAGGACCAATGTCA 1482
Db 317 GlyProGlyLysGluIleGlnProGlyCysGlnSer---LeuGlnArgAsnGlnCysAla 335
Qy 1483 TATTGTAAGAAAGAGGACACTGGGCAAGGAACCTGCCCAAG---AAGGGAACAAGGA 1539
Db 336 TyrCysLysGluIleGlyHisTrpLysAsnLysCysProGlnLeuLysGlyLysGlnGly 355
Qy 1540 CCAAGGATCTAGTCTCTAGAAGAAGATAAAGAT 1572
Db 356 Asp-----SerGluGlnGluAspProAsp 363
```

## RESULT 13

```
US-10-450-763-39289
; Sequence 39289, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
```

```
; SEQ ID NO 39289
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (325)..(341)
; OTHER INFORMATION: Eukaryotic and viral aspartyl proteases proteins domain
; OTHER INFORMATION: identified by eMATRIX, accession number BL00141A, p-value=4.150e-
; OTHER INFORMATION: 10, raw score of 12.10
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (9)..(214)
; OTHER INFORMATION: Gag p30 core shell protein domain identified by Pfam,
; OTHER INFORMATION: accession name Gag_p30, E-value=1.2e-67, Pfam score of 238.2
US-10-450-763-39289
```

```
Alignment Scores:
Pred. No.: 1.01e-30 Length: 688
Score: 526.50 Matches: 121
Percent Similarity: 53.9% Conservative: 50
Best Local Similarity: 38.2% Mismatches: 113
Query Match: 18.5% Indels: 33
DB: 5 Gaps: 9
```

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-10-450-763-39289 (1-688)

```
Qy 637 TATTGGCCCTTTTCTTCTGAGAGACTCTATTAATTGGAAACTAACCATCCCTTCTCTCG 696
Db 9 TyrGlnProPhePheSerAlaAspLeuLeuAsnTrpLysAsnAsnThrProSerTyrThr 28
Qy 697 GAGGATCCCCAACGCCCTCACGGGTTGGTGAGTCCCTTATGTTCTCTCACAGCCTACT 756
Db 29 GluLysProGlnAlaLeuIleAspLeuLeuGlnThrIleIleGlnThrHisAsnProThr 48
Qy 757 TGGGATGATTCTCAACAGCTGCTGCAGACACTTCTTCAACACGAGGACGAGAGAAATT 816
Db 49 TrpAlaAspCysHisGlnLeuLeuMetPheLeuPheAsnThrAspGluArgArgVal 68
Qy 817 CTATTAGAGGCTAGAAA-----AATGTTCTCGGGCGCGCGGGACCCACG 864
Db 69 LeuGlnAlaAlaThrLysTrpLeuGluGluHisAlaProAlaAspTyrGlnAsnPro--- 87
Qy 865 CGTTGCAAAATGAGATTGACATGGATTTCCTTAACTCGCCCGGTTGGGACTACAAC 924
Db 88 -----GlnGluTyrValArgThrGlnLeuProGlyThrAspProGlnTrpAspPro 105
Qy 925 ACGGCTGAAGGTAGGAGAGCTTGAAAAATCTATCGCCAGGCTCTGGTGGCGGCTCTCCG 984
Db 106 GluArgGluAspMetGlnArgLeuAsnArgTyrArgGluAlaLeuLeuGluGlyLeuLys 125
Qy 985 GCGGCTCAAGACGGGCCCACTAATTGGCTAAGGTAAAGAAAGTATGTCAGGAGCCGAAT 1044
Db 126 ArgGlyAlaGlnLysAlaThrAsnValAla-----IleGlnGlyLysGlu 140
Qy 1045 GAACCCCTCTGTTTTCTTGAGAGGCTCTTGGAGCCTTCAGCGGCTTACAGCCCTTTT 1104
Db 141 GluSerProAlaGlnPheTyrGluArgLeuCysGluAlaTyrCysMetTyrThrProPhe 160
Qy 1105 GATCCCACTCAGAGGCCCAAAAGCCCTCAGTGGCTTTGGCCTTTATAGGACAGTCAGCC 1164
Db 161 AspProAsnSerProGluAsnGlnArgMetIleAsnMetAlaLeuValSerGlnSerAla 180
Qy 1165 TTGGATATTAGAAAGCTTCAGAGACTGGAGGTTACAGAGGCTGAGGTAGTGTAGTGTAT 1224
Db 181 GluAspIleArgArgLysLeuGlnLysGlnAlaGlyPheAlaGlyMetAsnThrSerGln 200
Qy 1225 CTAGTGAAGGAGGAGAGAAAGTATTATACAAAGGAGACAGAGAAGAAAGGAAACAA 1284
Db 201 LeuLeuGluIleAlaAsnGlnValPheValAsnArgAspAlaValSerProLysGluAsn 220
Qy 1285 AGAAAGAGAGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1344
```

Db 109 ---LeuProProSerAlaProSerLeuProLeuGluProProLeuSerThrProProArg 121  
 Qy 418 -----GCTTGGCCGGAACCCCAA----- 435  
 Db 128 SerSerLeuTyrProAlaLeuThrProSerLeuGlyAlaLysProLysProGlnValLeu 147  
 Qy 436 -----TCTGTTCCCCACCCCTTATCTGTGCA 462  
 Db 148 SerAspSerGlyProLeuIleAspLeuLeuThrGluAspProProTyr----- 165  
 Qy 463 CAGGGTCCCGGAGGGGACCTTCTCCCTCCT----- 495  
 Db 166 -----ArgAspProArgProProSerAspArgAspGlyAsnGlyGlyGlu 181  
 Qy 496 GGAGCTCCGGCGGTGGAGGACCT-----GCTGCAGGGACTCGGAGCCGG 540  
 Db 182 AlaThrProAlaGlyGluAlaProAspProSerProMetAlaSerArgLeuArgGlyArg 201  
 Qy 541 AGGGCGCCACCCCGGAGCGGACAGACGAGATCGCGCATTTACCGCTCGCGACGTACGCG 600  
 Db 202 ArgGluProProValAlaAspSerThrThrSerGlnAlaPheProLeuArgThrGlyGly 221  
 Qy 601 CCTCCACACCGGGGGGCCAATTGCAGGCCCTCCAGTATTGGCCCTTTTCTTCTGCAGAT 560  
 Db 222 -----AsnGlyGln-----LeuGlnTyrTrpProPheSerSerSerAsp 234  
 Qy 661 CTCTATAATTGGAAACCTAAC-----CATCCCCCTTCTCTCG 696  
 Db 235 LeuTyrIleThrProValAsnSerLeuGluLysHisSerTrpTyrHisGlyProValSer 254  
 Qy 697 GAGGAT 702  
 Db 255 ArgAsn 256  
 RESULT 15  
 US-10-732-923-13455  
 ; Sequence 13455, Application US/10732923  
 ; Publication No. US20050108791A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edgerton, Michael D  
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
 ; FILE REFERENCE: 38-15(52796)C  
 ; CURRENT APPLICATION NUMBER: US/10/732,923  
 ; CURRENT FILING DATE: 2003-12-10  
 ; PRIOR APPLICATION NUMBER: 10/310,154  
 ; PRIOR FILING DATE: 2002-12-04  
 ; NUMBER OF SEQ ID NOS: 24149  
 ; SEQ ID NO 13455  
 ; LENGTH: 981  
 ; TYPE: PRT  
 ; ORGANISM: Abelson murine leukemia virus  
 US-10-732-923-13455  
 Alignment Scores:  
 Pred. No.: 1,8e-28 Length: 981  
 Score: 497.00 Matches: 116  
 Percent Similarity: 48.9% Conservative: 22  
 Best Local Similarity: 41.1% Mismatches: 70  
 Query Match: 17.5% Indels: 74  
 DB: 5 Gaps: 10  
 US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-10-732-923-13455 (1-981)  
 Qy 1 ATGGGACAGAGGTGACGACCCCTTAGTTGACTCTCGACCATTTGGACTGAAGTTAAA 60  
 Db 1 MetGlyGlnThrValThrThrProLeuSerLeuThrLeuGlyHisTrpLysAspValGlu 20  
 Qy 61 TCCAGGGCTCATATTTGTCTAGTTTCAGGTTAAGAGGACCTTGGCAGACTTCTGTGTC 120  
 Db 21 ArgIleAlaHisGlnSerValAspValLysLysArgGTrpValThrPheCysSer 40  
 Qy 121 TCTGAATGGCGCACATTGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCAGATT 180  
 Db 121 TCTGAATGGCGCACATTGATGTTGGATGGCCATCAGAGGGGACCTTTAATTCAGATT 180

```
Db      41  AlaGluTrpProThrPheAsnValGlyTrpProArgAspGlyThrPheAsnArgAspLeu  60
Qy      181 ATCTGGCTGTTAAAGCAGTTATTTTTCAGACTGGACCGGCTCTCATCCCGATCAGGAG  240
Db      61  IleThrGlnValLysIleLysValPheSerProGlyProHisGlyHisProAspGlnVal  80
Qy      241 CCCTATATCCTTACGTGGCAGATTGTGGCAGAGGATCTCCGCCATGGTTAAACCATGG  300
Db      81  ProTyrIleValThrTrpGluAlaLeuAlaPheAspProProProTrpValLysProPhe  100
Qy      301 CTGAATAAGCCAAAGACAGGTCCCGCAATTCTGGCTTTGGAGAGAAAAACAACAC  360
Db      101 ValHis----ProLysProProPro-----  108
Qy      361 TCGGCTGAAAAGTCAAGCCCTCTCCTCATATCTACCCCGAGATTGAGGAGCCACCG---  417
Db      109 ---LeuProProSerAlaProSerLeuProLeuGluProProLeuSerThrProProArg  127
Qy      418 -----GCTTGGCCGGAACCCCAA-----  435
Db      128 SerSerLeuTyrProAlaLeuThrProSerLeuGlyAlaLysProLysProGlnValLeu  147
Qy      436 -----TCTGTTCCCCACCCCTTATCTGGCA  462
Db      148 SerAspSerGlyGlyProLeuIleAspLeuLeuThrGluAspProProTyr-----  165
Qy      463 CAGGTGCCGAGGGGACCCCTTTCCTCCT-----  495
Db      166 -----ArgAspProArgProProProSerAspArgAspGlyAsnGlyGlyGlu  181
Qy      496 GGAGCTCCGGCGGTGGAGGACCT-----GCTGCAGGGACTCGGAGCCGG  540
Db      182 AlaThrProAlaGlyGluAlaProAspProSerProMetAlaSerArgLeuArgGlyArg  201
Qy      541 AGGGCGCCACCCCGAGCGGACAGACGAGATCGGCACATTACCGCTCGGCACGTACGCGC  600
Db      202 ArgGluProProValAlaAspSerThrThrSerGlnAlaPheProLeuArgThrGlyGly  221
Qy      601 CCTCCACACCGGGGGCCAAATTGCGCCCTCCAGTATTGGCCCTTTTCTCTGCAGAT  660
Db      222 -----AsnGlyGln-----LeuGlnTyrTrpProPheSerSerSerAsp  234
Qy      661 CTCTATAATTGGAATAACTAAC-----CATCCCCCTTTCTCG  696
Db      235 LeuTyrIleThrProValAsnSerLeuGluLysHisSerTrpTyrHisGlyProValSer  254
Qy      697 GAGGAT 702
Db      255 ArgAsn 256
```

Search completed: February 14, 2006, 17:17:27  
Job time : 223.636 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 16:19:01 ; Search time 1.37849 Seconds  
(without alignments)  
2992.933 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_585\_2156

Perfect score: 2848

Sequence: 1 ATGGGACAGCGGTGACGAC.....CTCTAGAGAGATAAAGAT 1572

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 194028

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abs5/ABSSWEB\_spool/US10723552/runat 14022006 125151 13368/app query.fasta\_1  
-DB=Published\_Applications\_AA\_New -QFMT=fastan -SURFIX=rapbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=bl0sum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-MAXLEN=2000000000 -HOST=abs02p  
-USR=US10723552\_@CNG 1 1 17 @runat 14022006 125151 13368 -NCPU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THRAUS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-WGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA New:

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1358	47.7	537	7	US-11-129-442-47
2	163.5	5.7	416	7	US-11-226-657-136
3	138.5	4.9	1388	6	US-10-821-234-1143
4	136.5	4.8	538	7	US-11-124-368A-311
5	136.5	4.8	538	7	US-11-124-368A-312
6	136.5	4.8	538	7	US-11-124-368A-313
7	136.5	4.8	1560	7	US-11-059-982-1
8	132	4.6	1410	6	US-10-821-234-1050
9	129.5	4.5	2760	7	US-11-124-367A-444
					Sequence 47, Appl
					Sequence 136, App
					Sequence 1143, Ap
					Sequence 311, App
					Sequence 312, App
					Sequence 313, App
					Sequence 1, Appli
					Sequence 1050, Ap
					Sequence 444, App

10	129.5	4.5	2803	7	US-11-124-367A-442	Sequence 442, App
11	129.5	4.5	2803	7	US-11-124-367A-445	Sequence 445, App
12	129.5	4.5	2984	7	US-11-124-367A-443	Sequence 443, App
13	129.5	4.5	3027	7	US-11-124-367A-441	Sequence 441, App
14	129	4.5	1744	7	US-11-182-016-22	Sequence 22, Appl
15	128.5	4.5	615	6	US-10-982-545-14	Sequence 14, Appl
16	128	4.5	948	6	US-10-523-477-14	Sequence 14, Appl
17	127.5	4.5	716	7	US-11-150-845-16	Sequence 16, Appl
18	127.5	4.5	716	7	US-11-150-847-16	Sequence 16, Appl
19	127	4.5	638	7	US-11-150-845-24	Sequence 24, Appl
20	127	4.5	638	7	US-11-150-847-24	Sequence 24, Appl
21	127	4.5	689	7	US-11-024-959-510	Sequence 510, App
22	125.5	4.4	616	6	US-10-982-545-5	Sequence 5, Appli
23	123	4.3	639	6	US-10-821-234-907	Sequence 907, App
24	123	4.3	863	7	US-11-169-041-167	Sequence 167, App
25	123	4.3	915	6	US-10-821-234-1514	Sequence 1514, Ap
26	121.5	4.3	8746	7	US-11-098-686-10212	Sequence 10232, A
27	121	4.2	537	6	US-10-504-364-6	Sequence 6, Appli
28	121	4.2	537	6	US-10-504-364-7	Sequence 7, Appli
29	120.5	4.2	551	6	US-10-821-234-1580	Sequence 1580, Ap
30	120.5	4.2	551	6	US-10-504-364-1	Sequence 1, Appli
31	120.5	4.2	551	6	US-10-504-364-2	Sequence 2, Appli
32	120.5	4.2	551	6	US-10-504-364-5	Sequence 5, Appli
33	119.5	4.2	715	7	US-11-150-845-14	Sequence 14, Appl
34	119.5	4.2	715	7	US-11-150-847-14	Sequence 14, Appl
35	119.5	4.2	1516	6	US-10-220-824-8	Sequence 8, Appli
36	119.5	4.2	1823	6	US-10-995-561-988	Sequence 988, App
37	119.5	4.2	2157	6	US-10-995-561-991	Sequence 991, App
38	119	4.2	609	7	US-11-150-845-20	Sequence 20, Appl
39	119	4.2	609	7	US-11-150-847-20	Sequence 20, Appl
40	119	4.2	891	7	US-11-182-016-38	Sequence 38, Appl
41	119	4.2	2102	6	US-10-995-561-990	Sequence 990, App
42	119	4.2	2108	6	US-10-995-561-989	Sequence 989, App
43	117.5	4.1	915	6	US-10-995-561-1003	Sequence 1003, Ap
44	117.5	4.1	917	6	US-10-995-561-1000	Sequence 1000, Ap
45	117.5	4.1	940	6	US-10-995-561-1004	Sequence 1004, Ap

ALIGNMENTS

RESULT 1

US-11-129-442-47

; Sequence 47, Application US/11129442  
; Publication No. US2006002951A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleiman, Lawrence  
; APPLICANT: CEN, Shan  
; APPLICANT: GUO, Fei  
; TITLE OF INVENTION: Inhibition of the tRNAlys3-Primed Initiation of Reverse  
; FILE REFERENCE: 11168.257  
; CURRENT APPLICATION NUMBER: US/11/129,442  
; PRIOR FILING DATE: 2005-05-16  
; PRIOR APPLICATION NUMBER: CA 2,467,312  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 47  
; LENGTH: 537  
; TYPE: PRT  
; ORGANISM: Murine leukemia virus  
US-11-129-442-47

Alignment Scores:

Pred. No.: 3.92e-88 Length: 537  
Score: 1358.00 Matches: 277  
Percent Similarity: 62.7% Conservative: 76  
Best Local Similarity: 49.2% Mismatches: 142  
Query Match: 47.7% Indels: 68  
DB: 7 Gaps: 12

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-11-129-442-47 (1-537)



```
QY 1 ATGGACAGCGGTACGACCCCTCTAGTTTGACTCTCGACCATTTGACTGAAGTTAAA 60
Db |||||
QY 1 MetGlyGlnThrValThrProLeuSerLeuThrLeuGluHisTrpGlyAspValGln 20
Db |||||
QY 61 TCCAGGGCTCAATAATTTGTCAGTTTCAAGTTAAGAGGACCTTGGCAGACATTTCTGTGTC 120
Db |||||
QY 21 ArgIleAlaSerAsnGlnSerValAspValLysLysArgTrpValThrPheCysSer 40
Db |||||
QY 121 TCTGAATGCCGACATTCATGTTGGATGGCCATCAGAGGGACCTTTAAATCTGAGATT 180
Db |||||
QY 41 AlaGluTrpProThrPheGlyValGlyTrpProGlnAspGlyThrPheAsnLeuAspIle 60
Db |||||
QY 181 ATCTGGCTGTTAAAGCAGTTATTTTTCAGACTGCACCGCGCTCTCATCCGATCAGGAG 240
Db |||||
QY 61 IleLeuGlnValLysSerLysValPheSerProGlyProHisGlyHisProAspGlnVal 80
Db |||||
QY 241 CCTATATCTTACCTGGTCAAGATTGGCAGAGGATCTCCGCCATGGTGTAAACCATGG 300
Db |||||
QY 81 ProTyrIleValThrTrpGluAlaIleAlaTyrGluProProTrpValLysProPhe 100
Db |||||
QY 301 CTGAATAAGCCAGAAAGCCAGGTCCCGA-----ATTCTGGCTCTTGGAGAGAAAAC 354
Db |||||
QY 101 ValSerProLysLeuSerProSerProThrGlyProIleLeuProSerGlyProSerThr 120
Db |||||
QY 355 AAACACTCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCCGAGATTGAGGAGCCA 414
Db |||||
QY 121 GlnProProProArgSerAlaLeu---TyrProAlaLeuThrProSerIleLysProArg 139
Db |||||
QY 415 CCGGCTTGGCGGACCCCAATCT-----GTT 441
Db |||||
QY 140 ProSerLysProGlnValLeuSerAspAsnGlyGlyProLeuIleAspLeuLeuThrGlu 159
Db |||||
QY 442 CCCCACCCCTTATCTGGCACAGGTGCCGAGGGACCTTTGCCCTCTCTGGA--- 498
Db |||||
QY 160 AspProProProTyr-----GlyGluGlnGlyProSerSerSerAspGlyAsp 175
Db |||||
QY 499 -----GCTCCGGCGTGGAGGACCTCTGCA 525
Db |||||
QY 176 GlyAspArgGluGluAlaThrSerThrSerGluIleProAlaProSerProMetValSer 195
Db |||||
QY 526 GGGACTCGGAGCGGAGGGGCCACCCCGGAGCGGACAGACAGATCGGACATTACCG 585
Db |||||
QY 196 ArgLeuArgGlyLysArgAspProProAlaAlaAspSerThrSerArgValPhePro 215
Db |||||
QY 586 CTGCGCAGTACGGCCCTCCACACCGGGGGGCCAATTGCGAGCCCTCCAGTATTTGGCCC 645
Db |||||
QY 216 LeuArgLeuGlyGly-----AsnGlyGln-----LeuGlnTyrTrpPro 228
Db |||||
QY 646 TTTTCTCTCGAGATCTCTATAATTGGAATACTAACCATCCCTTTCTCGGAGGATCCC 705
Db |||||
QY 229 PheSerSerSerAspLeuTyrAsnTrpLysAsnAsnProSerPheSerGluAspPro 248
Db |||||
QY 706 CAACGCCCTCAGCGGCTTGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGATGAT 765
Db |||||
QY 249 GlyLysLeuThrAlaLeuIleGluSerValLeuThrThrHisGlnProThrTrpAspAsp 268
Db |||||
QY 766 TGTCAACAGCTGCTCAGACACTTTCACACCGGAGGCGAGAGAGAAATCTATTAGAG 825
Db |||||
QY 269 CysGlnGlnLeuLeuGlyThrLeuLeuThrGlyGlyGluLysGlnArgValLeuLeuGlu 288
Db |||||
QY 826 GCTGAAAAAATGTTCTTGGGCGCCAGCGGCGGACCCAGCGGTGCAAAATGAGATTGAC 885
Db |||||
QY 289 AlaArgLysAlaValArgGlyAsnAspGlyArgProThrGlnLeuProAsnGluValAsn 308
Db |||||
QY 886 ATGGATTTCCTTAACCTCGCCCGGTGGGACTACAAACAGGCTGGAAGGTAGGGAGAGC 945
Db |||||
QY 309 SerAlaPheProLeuGluArgProAspTrpAspTyrThrProGluGlyArgAsnHis 328
Db |||||
QY 946 TTGAAAAATCTATCGCAGGCTCTGTGGGGGTCTCCGGGGCGCTCAAGACGGCCCACT 1005
Db |||||
QY 329 LeuValLeuTyrArgGlnLeuLeuAlaGlyLeuGlnAsnAlaGlyArgSerProThr 348
Db |||||
QY 1006 AATTGGCTAAGTAAGAGAGTGTATGAGGGAGCCGAATGAACCCCTCTGTTTCTT 1065
Db |||||
```

```
Db |||||
QY 349 AsnLeuAlaLysValLysGlyIleThrGlnGlyProAsnGluSerProSerAlaPheLeu 368
Db |||||
QY 1066 GAGAGGCTCTTTGGAGCCTTCAGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAA 1125
Db |||||
QY 369 GluArgLeuLysGluAlaTyrArgArgTyrThrProTyrAspProGluAspProGlyGln 388
Db |||||
QY 1126 AAAGCCTCAGTGGCTTGGCCCTTTATAGGACAGTCAGCCTTGGATATTAGAAAGAGCTT 1185
Db |||||
QY 389 GluThrAsnValSerMetSerPheIleTrpGlnSerAlaProAspIleGlyArgLysLeu 408
Db |||||
QY 1186 CAGAGCTGGAAGGTTTACAGAGGCTGAGTTACGTGATCTAGTGAAGGAGCGCAGAGAAA 1245
Db |||||
QY 409 GluArgLeuGluAspLeuLysSerLysThrLeuGlyAspLeuValArgGluAlaGluArg 428
Db |||||
QY 1246 GTATATTACAAAAGGAGACAGAAAGGAAAGGAAACAAAGAGAGAGAGAGAGAGAG 1305
Db |||||
QY 429 IlePheAsnLysGlyGluThrProGluGluArgGluGluArgValArgArgGluThrGlu 448
Db |||||
QY 1306 CAAAGGAGGAGAAACA-----CGTAATAAACGG 1332
Db |||||
QY 449 GluLysGluGluArgArgAlaGluGluGlnLysGluLysGluLysGluLysArgArg 468
Db |||||
QY 1333 CAAGAGAAGAAATTTGACTTAAGATCTTGCTCGAGTGTGAAGGGAAGCAATACGCAA 1392
Db |||||
QY 469 ArgHisArgGluMetSerLysLeuAlaThrValValSerGlyGln----- 484
Db |||||
QY 1393 AGAGAGAGAGATTTTAGGAAATTTAGTTCAGCCCTAGACAGTCAGGGAACTGGGCAAT 1452
Db |||||
QY 485 -----ArgGlnAspArgGlnGlyGluArg-----Arg 494
Db |||||
QY 1453 AGGACCCACTCGACAGGACCAATGTGCATTTGTAAGAAAGAGGACACTGGGCAAGG 1512
Db |||||
QY 495 ArgProGlnLeuAspLysAspGlnCysAlaTyrCysLysGluLysGlyHisTrpAlaLys 514
Db |||||
QY 1513 AACTGCCCAAGAAG-----GGAAACAAAGCAACCAAG-----ATCCTAGCT 1554
Db |||||
QY 515 AspCysProLysLysProArgGlyProArgGlyProArgGlyProGlnThrSerLeuLeuThr 534
Db |||||
QY 1555 CTAGAAGAA 1563
Db |||||
QY 535 LeuAspAsp 537
Db |||||

RESULT 2
US-11-226-657-136
; Sequence 136, Application US/11226657
; Publication No. US20060025574A1
; GENERAL INFORMATION:
; APPLICANT: Steven M. Ruben, et al.
; TITLE OF INVENTION: 32 Human Secreted Proteins
; FILE REFERENCE: P2006G13APIC1D2
; CURRENT APPLICATION NUMBER: US/11/226,657
; CURRENT FILING DATE: 2005-09-15
; PRIOR APPLICATION NUMBER: 10/062,831
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 09/690,454
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: 09/189,144
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/US98/10868
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/044,039
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,093
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,190
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/050,935
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,101
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,356
; PRIOR FILING DATE: 1997-05-30
```



```
Db      213 GlnIleGlnArgHisAsnGlnSerThrAlaIleAsnLeuAsnAsnProGluSerGln 232
Qy      200 TTATTTTCAGACTGGACCGGCTCTCATCCGATCAGAGCCCTATATCTTACGTGGC 259
Db      233 ---SerMetHisLeuGluThrArgLeuValGlnLeuAspSerAlaIleSerMet----- 249
Qy      260 AAGATTTCGCGAGAGGATCTCCGC-----CATGGTTAAACCATGGCTGA 304
Db      250 GluLeuTrpGlnGluAlaPheLysAlaValGluAspIleHisGlyLeu----- 265
Qy      305 ATAAGCCAAAGAACCGAGTCCCGCAATCTGGCTCTTGGAG-----AGAAAAACA 355
Db      266 PheSerLeuSerLysLeuProLysProGlnLeuMetAlaAsnTyrAsnLysVal 285
Qy      356 AACACTCGGCTGAAAAAGTCAAGCCCTCTCTCATATCTACCCGAGATTGAGAGCCAC 415
Db      286 SerThrValPheTrpLysSerGlyAsnAlaLeuPheHisAlaSerThrLeu-----His 303
Qy      416 CGGCTTGGCCGGAACCCCAATCTGTTCCTCCACCCCTTATCTGCACAGGTCGCGGA 475
Db      304 ArgLeu-----TyrHisLeuSerArg 310
Qy      476 GGGACCCCTTGGCCCTCTCGAGCTCCGGGG-----TGGAGGGACCTGTCGAGGA 529
Db      311 -----GluMetArgLysAsnLeuThrGlnAspGluMetGlnArg 323
Qy      530 CTCGAGCGCGGAGG-----GGCGCCACCCCGGAGGCGGACAG 565
Db      324 Met--SerThrArgValLeuLeuAlaThrLeuSerIleProIleThrProGluArgThrA 343
Qy      566 AC-----GAGATCGCGA 577
Db      343 spIleAlaArgLeuLeuAspMetAspGlyIleIleValGluLysGlnArgGluAlaTr 363
Qy      578 CATTACCCTCGCGACGTACGCCCTCCACACCGGGGGC-----C 619
Db      363 hrLeu---LeuGlyLeuGlnAlaProThrArgIleGlyLeuIleAsnAspMetValA 382
Qy      620 AATTCGACCCCTCCAGTATTGGCCCTTTCTTCTGCGAGATCTCTATATTTGG---AAAA 676
Db      382 rgPheAsnValLeuGlnTrpValValProGluValLysAspLeuTrpAsnTrpLeuGluV 402
Qy      677 CTACCATCCCTCTTCTCG-----G 697
Db      402 alGluPheAsnProLeuLysLeuCysGluArgValThrLysValLeuAsnTrpValArgG 422
Qy      698 AGGATCC----- 705
Db      422 luGlnProGluLysGluProGluLeuGlnGlnTrpValProGlnLeuGlnAsnAsnThrI 442
Qy      706 -----CAACGCTCAGCGGTGGTGGAGTCCCTTATGTCTCTACACGAC 751
Db      442 leLeuArgLeuGlnValSerGlnIleTrpGlnSerIleGluPheSerArgLeuT 462
Qy      752 CTACT-----TGGATGATTTCTACAGCTCTCTGACACACTCTTCACACCGGAGG 802
Db      462 hrSerLeuValProPheValAspAlaPheGlnLeu----- 473
Qy      803 AGCGAGAGAGATTTCTATTAGAGGCTAGAAAAAATGTTCTTGGGGCCGACGGCGACCCA 862
Db      474 -----GluArgAlaIleValAspAla-----AlaArgHisC 484
Qy      863 CGCGGTTGCAAAATGAGATTGACATGGGATTTCCCTTAAGTTCGCCCGCGTGGGACTACA 922
Db      484 ysAspLeuGlnValArgIleAspHisThrSerArgThrLeuSerPheGlySerAspLeuA 504
Qy      923 ACACGGCTCAGAGTAGGAGGAGCTTGAATAATCTATCGCAGGCTCTGGTGGCGGTCTCC 982
Db      504 snTyrAlaThr---ArgGluAspAlaProIle----- 513
Qy      983 GGGGCGCCTCAAGACGGGCCACTAATTTGGCTAAGGTAAGAGAGTGTGCGGAGCGGA 1042
Db      514 -----GlyProH 516
```

```
Qy      1043 ATGAACCCCTCTGTGTTTTCTTGAGAGGCTCTTGAAGCCTTCAGCGGTACACCCCTT 1102
Db      516 is-----LeuGlnSerMet----- 520
Qy      1103 TTGATCCACCTCAGAGGCCCAAAAAGCCTCAGTGGCTTTGGCCTTTATAGGACAGTCAG 1162
Db      521 -----ProSerGluGlnIleArgAsnGlnLeuThrAlaMetSerValLeuAlaLysA 539
Qy      1163 CTTGGATATTAGAAAGCTTCAGACTTGGAGGGTTACAGAG----- 1209
Db      539 laLeuGluValIleLysProAlaHisIleLeuGlnGluLysGluGlnHisGlnLeuA 559
Qy      1210 -----GCTGACTTACGTCTAGTGAAGGAGGACAGAAAGTATATTACAAAAGGG 1261
Db      559 laValThrAlaTyrLeuLysAsnSerArgLysGluHisGlnArgIleLeuAlaArgG 579
Qy      1262 AGACAGAAGAAAGGAAACAAAAGAG-----AGAAAAGAGAGG 1306
Db      579 lnThrIleGluLysGluArgLeuSerLeuAsnIleGlnArgGluLysGluG 599
Qy      1307 AAAGGAGGAAGA-----C 1321
Db      599 luLeuGluGlnArgGluAlaGluLeuGlnLysValArgLysAlaGluGluArgLeuA 619
Qy      1322 GTAATAACGCAAGAGAGAAATTTGACTAGATCTTGGCTGCAGTGGTTGAAGGAAAA 1381
Db      619 rgGlnGluAlaLysGluArgLysGluArgIleLeuGlnGluHisGlnIleLysL 639
Qy      1382 GCAATACGAAAGAGAGAGA 1401
Db      639 ysLysThrValArgGluArg 645
```

## RESULT 4

```
US-11-124-368A-311
; Sequence 311, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 2112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 311
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-368A-311
```

## Alignment Scores:

Pred. No.:	0.0123	Length:	538
Score:	136.50	Matches:	83
Percent Similarity:	31.7%	Conservative:	23
Best Local Similarity:	24.9%	Mismatches:	112
Query Match:	4.8%	Indels:	116
DB:	7	Gaps:	12

US-10-723-552-3\_copy\_585\_2156 (1-1572) x US-11-124-368A-311 (1-538)

```
Qy      301 CTGATAAGCCCAAGAAAGCCAGTCCCGCAATTCGGCTCTTGG----- 345
Db      41 LeuGlyGlyProArgLysSerGlyProArgGlyAlaThrGlyGlyProGlyAspGluPro 60
Qy      346 -----GAGAAAAACAAACACTCGGCTGAAAAAGTCAAGCCCTCTCCTCAT 390
```





US-11-059-982-1

Alignment Scores:

Pred. No.: 0.0131 Length: 1560  
Score: 136.50 Matches: 95  
Percent Similarity: 29.4% Conservative: 38  
Best Local Similarity: 21.0% Mismatches: 156  
Query Match: 4.8% Indels: 163  
DB: 7 Gaps: 20

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-11-059-982-1 (1-1560)

Qy 151 CCATCAGAGGGGACCTTAATCT-----GAGATTATCTCGGCTGTTAAAGCAGTTATT 204  
Db 666 ProSerProGlyLeuAlaSerSerProGlyLysValLeu----- 679  
Qy 205 TTTCCAGACTGACCGGCTCTCATCCGATCAGGAGCCCTATATCTTACGTGCAAGAT 264  
Db 680 -----GlyGlnProSerAlaThrProThrAlaIleLeuThrGlnAsp 694  
Qy 264 ----- 264  
Db 695 SerLeuGlnMetPheLeuProGlnGluArgSerGlnGlnProLeuSerAlaGluGlyPro 714  
Qy 265 -----TTGCAGAGGATCTCCGCCATCTCGCCATGGGTAAACCATGG 300  
Db 715 HisLeuSerValProAlaSerValIleValSerAlaProProProAlaGlnAspPro--- 733  
Qy 301 CTGAATAAGCCCAAGAACCCAGGTCCTCGGATTTCTGGCTCTTGGAGAGAGAAAAACACAC 360  
Db 734 -----AlaProAlaThrProValAlaLysGlyAlaGlyLeuGlyProGlnAlaProAsp 751  
Qy 361 TCGGCTGAAAAGTCAAGCCCTCTCTCATATC----- 393  
Db 752 SerGlnAlaSerProAlaProAlaProGlnIleProAlaAlaAlaProLeuLysGlyPro 771  
Qy 394 -----TACCCCGAGATTGAGAGCCACCGGCTTGG 423  
Db 772 GlyProSerSerProSerLeuProHisGlnAlaProLeuGlyAspSerProHisLeu 791  
Qy 424 CCGGAACCCCAATCTGTTCGCCA-----CCCCCTATCTGGCAGAGGTGCGCG 474  
Db 792 ProSerProHisProThrArgProProSerArgProProSerArgProGlnSerValSer 811  
Qy 475 AGGCGACCTTTGCGCCCTCTCTGGAGCTCCGCGGTGGAGGACCTGCTGCGAGGACT--- 531  
Db 812 ArgProProSerGluProProLeuHisProCysProProProGlnAlaProProThrLeu 831  
Qy 532 -----CGAGCCGAGGGGCGCCACCCCGGAGCGGACAGCAGATC 573  
Db 832 ProGlyIlePheValIleGlnAsnGlnLeuGlyValProProAlaSerAsnProAla 851  
Qy 574 GCGACATTACGCTGCGCAGTACGCGCTCCC----- 606  
Db 852 ProThrAlaPro-----GlyProProGlnProProLeuArgProGlnSerGln 867  
Qy 607 ACACCGGGGGCCAAATTCAGCCCTCCAGTATGGCCCTTTCTCTGCGAGATCTC--- 663  
Db 868 ProProGluGlyProLeuProAlaProHisLeuProProSerSerThrSerSerAla 887  
Qy 664 -----TATAATTGGAAA 675  
Db 888 ValAlaSerSerSerGluThrSerSerArgLeuProAlaProThrProSerAspPheGln 907  
Qy 676 ACTAACCATCCCCCTTTCTCGGAGGATCCCAACGCTCAGGGGTGGTGGAGTCCCTT 735  
Db 908 LeuGlnPheProPro---SerGlnGlyProHisLysSerProThrProProThrLeu 926  
Qy 736 ATGTTCTCTCACACGCT-----ACTTGGGATGATTGT 768  
Db 927 HisLeuValProGluProAlaAlaProProProProProProProProProProPhe----- 943  
Qy 769 CAACAGCTGTGCAGACACTTTTCAACACCGAGGAGCGAGAGAAATTTCTATTAGAGGCT 828

Db 944 ---GlnMetValThrThrProPheProAlaLeuProGlnProLysAlaLeuLeuGluArg 962  
Qy 829 AGAAAAAATGTTCTCTGGG-----GCCGACGGCGACCCACGCGG 867  
Db 963 PheHisGlnValProSerGlyIleIleLeuGlnAsnLysAlaGlyGlyAlaProAlaAla 982  
Qy 868 TTGCAAAATGAGATTGACATCGGGATTTCCCTTAACTCGCCCGGTTGGGACTACACACG 927  
Db 983 ProGlnThrSerThrSerLeuGly---ProLeuThrSerProAla----- 996  
Qy 928 GCTGAGGTAGGAGAGCTTGAAAATCTATCGCCAGGCT----- 966  
Db 997 -----AlaSerValLeuValSerGlyGlnAlaProSerGlyThrProThrAla 1012  
Qy 967 -----CTGGTGGCGGCTCTC----- 981  
Db 1013 ProSerHisAlaProAlaProAlaProMetAlaAlaThrGlyLeuProProLeuLeuPro 1032  
Qy 982 -----CGGGCGGCTCAAGACGCGCCACT---AATTTGGCTAAGTAAGAA 1026  
Db 1033 AlaGluAsnLysAlaPheAlaSerAsnLeuProThrLeuAsnValAlaLysAla----- 1050  
Qy 1027 GTGATGAGGAGACCGAATGAACCCCTCTGTGTTTCTTGAGAGGCTCTTGGAAGCCTTC 1086  
Db 1051 AlaSerSerGlyProGlyLysProSerGlyLeuGlnTyrGluSerLysLeuSerGlyLeu 1070  
Qy 1087 AGGCGGTACACCCCTTTTGATCCACCTCAGAGGCC 1122  
Db 1071 LysLysProProThrLeuGlnProSerLysGluAla 1082

RESULT 8  
US-10-821-234-1050  
; Sequence 1050, Application US/10821234  
; Publication No. US20050255114A1  
; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Andarmani, Susan  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia  
; FILE REFERENCE: 821A  
; CURRENT APPLICATION NUMBER: US/10/821,234  
; PRIOR FILING DATE: 2004-04-07  
; PRIOR APPLICATION NUMBER: US 60/462,047  
; NUMBER OF SEQ ID NOS: 1704  
; SOFTWARE: pt\_seq\_genes version 1.0  
; SEQ ID NO 1050  
; LENGTH: 1410  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-821-234-1050

Alignment Scores:  
Pred. No.: 0.0269 Length: 1410  
Score: 132.00 Matches: 84  
Percent Similarity: 31.9% Conservative: 58  
Best Local Similarity: 18.9% Mismatches: 165  
Query Match: 4.6% Indels: 138  
DB: 6 Gaps: 16

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-10-821-234-1050 (1-1410)

Qy 346 GAGAAAAACAAACATCGGCTGAAAAGTCAAGCCCTCTCTCATATCTACCCGAGATT 405  
Db 986 GluHisLeuAsnHisSerLysSerValThrProAlaSerThrLeuThrLys----- 1003  
Qy 406 GAGGAGCCACCGCTTGGCGGAAACCCCATCTGTT----- 441  
Db 1004 SerGlyProGlyArgTrpLysThrProAlaAlaIleProAlaThrProValAlaValSer 1023  
Qy 442 -----CCCCCACCCCTTATCTGCGACAGGGT 468

```
Db      1024 GlnProIleArgThrAspLeuProProProProProProValHisTyAlaGly 1043
Qy      469 CCGCGGAGGGA-----CCCTTGGCCCTCT----- 495
Db      1044 AspPheAspGlyMetSerMetAspLeuProProProProSerAlaAsnGlnIle 1063
Qy      496 GGAGCTCCGGCGGTGGAGGACCTGCTCAGGAGCTCGGAGCCG----- 540
Db      1064 GlyLeuProSerAlaGlnValAlaAlaGluArgArgLysArgGluGluHisGlnArg 1083
Qy      541 -----AGGGCGCCACCCCGGAGCGGACGAC 567
Db      1084 TrpTyrgluLysGluLysAlaArgLeuGluGluArgGluArgLysArgGluGln 1103
Qy      568 GAGATCGGCATATACCGCTCGGACGTACGGCCCTCCACACCGGGGGGCAATTGGAG 627
Db      1104 GluArgLysLeuGlyGlnMetArgThr-----Gln 1113
Qy      628 CCCCTCCAGTATGGCCCTTTCTTCTGCAGATCTCTATATTGAAACTAACCATCCC 687
Db      1114 SerLeuAsnProAlaProPheSerProLeuThrAlaGlnGlnMetLysProGluLysPro 1133
Qy      688 CTTTCTCGAGGATCCCAACCGCTCAGGGGTTGGTGAGTCCCTTATTTCTCTCAC 747
Db      1134 SerThrLeuGlnArgProGlnGluThr-----ValIleArgGluLeuGlnProGlnGln 1151
Qy      748 CAGCTACTCTGGGATGATGTCAACAGCTGTGCAGACACTTTCACAAACGAGAGCGGA 807
Db      1152 GlnProArg-----ThrIleGluArgArg 1159
Qy      808 GAGAGATTCTATTAGCGGTAGAAAAATGTTCTCTGGGCGGACGGCGACCGCG 867
Db      1160 AspLeuGlnTyrlleThrValSerLys----- 1168
Qy      868 TTGCAAAATGAGATTGACATGGGATTTCCCTTAACCTCGCCCGTTGGGACTACAACACG 927
Db      1169 -----GluGluLeuSerSerGlyThrValCysProProAspProTrp-----Lys 1183
Qy      928 GCTGAGGTAGGAGAGCTTGAATCTATCGCCAG-----GCTCTGGTG 972
Db      1184 ArgAlaLysGluLysGlnGlnMetHisIleValAspMetLeuSer 1203
Qy      973 CGGGTCTCCGGGGCGCTCAAGCGGCCACTAATTTGGCT-----AAGTA 1020
Db      1204 LysGluIleGlnGluLeuGlnSerLysProAspArgSerAlaGluLysSerAspArgLeu 1223
Qy      1021 AGAGAAGTGATG----- 1032
Db      1224 ArgLysLeuMetLeuGluTrpGlnPheGlnLysArgLeuGlnGlnSerLysGlnLysAsp 1243
Qy      1033 -----CAGGACCGNATGAACCCCTCTGTTTTCTTGGAGGCTCTTG 1077
Db      1244 GluAspAspGluGluGluAspAspValAspThrMetLeuIleMetGlnArgLeu 1263
Qy      1078 GAAGCTTCAGCGGTACACCCCTTTGATCCACCTCAGAGGCCCAAAAGCCCTCAGTG 1137
Db      1264 GluAlaGluArgArg-----AlaArgThrAlaMetPro 1274
Qy      1138 GCTTTGGCTTTATAGGACAGTCAGCCTTGGATATTAGAAAGAGCTTCAGAGACTGGAA 1197
Db      1275 AlalleSerValLeuAspLeuGlnAspGluGluArgArgGlnGlnGlnGlu 1294
Qy      1198 GGGTTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGAGGAGAAAGTATTACAAA 1257
Db      1295 GluMetArgLysArgGluAlaGluAspArgAlaArgGlnGluGlu-----Arg 1311
Qy      1258 AGGGAGACAGAGAGAAAGGAAACAAAGAAAGAGAGAGAGAGAGAAAGGAGGAA 1317
Db      1312 ArgArgGlnGluGluGluArgThrLysArgAspAlaGluGluLysArgArgGlnGlu 1331
Qy      1318 -----AGCGTAAATAACGGCAAGAGAAAGATTTGACTAAG 1353
```

```
Db      1332 GlyTyrtyrSerArgLeuGluAlaGluArgArgGlnHisAspGluAlaAlaArgArg 1351
Qy      1354 ATCTTGGCTCAGTGGTTGAGGAGAAAAGCAATACGGAAGAGAGAGATTTAGGAAA 1413
Db      1352 LeuLeuGluProGluAlaProGlyLeuCysArgProProLeuProA-AspTyTyrGluPro 1371
Qy      1414 ATTAGGTCCAGGCCCT 1428
Db      1372 ProSerProSerPro 1376

RESULT 9
US-11-124-367A-444
; Sequence 444, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 444
; LENGTH: 2760
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-367A-444

Alignment Scores:
Pred. No.: 0.0419 Length: 2760
Score: 129.50 Matches: 71
Percent Similarity: 42.3% Conservative: 72
Best Local Similarity: 21.0% Mismatches: 119
Query Match: 4.5% Indels: 76
DB: 7 Gaps: 16

US-10-723-552-3_COPY_585_2156 (1-1572) x US-11-124-367A-444 (1-2760)
Qy      712 CTCACGGGGTGTGGAGTCCCTTATGTTCTCTCACAGCCTACTTGGGATGATTGTCAA 771
Db      247 IleThrAlaLeuVal-----ValTrpLeuProAlaAsnProThr-----Glu 260
Qy      772 CAGCTGTGCAGACACTCTTCACAAACCGAGGAGGAGAGAAATTTCTATTAGAGGCTAGA 831
Db      261 LysIleValArgValLeuPheProGlyAsnAlaProGlnAsnLysIleLeuGluGlyLeu 280
Qy      832 AAAAATGTTCTCGGGCGGAC-----GGCGAGCCCGCGGTTGCAAAATGAGATTGAC 885
Db      281 GluLysLeuArgHisLeuAspPheLeuArgTytrProValAlaThrGlnLysAspLeuAla 300
Qy      886 ATGGGATTTCCCTTAAC-----CGCCCCGGTTGGGACTACACACCGCTGAAGGTAGG 939
Db      301 SerGlyAlaValProThrAsnLeuLysProSerLysIleLysGlnArgAlaAspSerLys 320
Qy      940 GAGAGCTTGAATCTATCGCCAGGCTCTGGTGGCGGCTC----- 981
Db      321 GluSerLeuLysAlaThrThrLysThrAlaValSerLysLeuAlaLysArgGluGluVal 340
Qy      982 -----CGGGCGGCTCAAGACCGCCCACTAATTTGGCTAAG-----GTA 1020
Db      341 ValGluGluGlyAlaLysGluAlaArgSerGluLeuAlaLysGluLeuAlaLysThrGlu 360
Qy      1021 AGAAGATGATGAGGAGCGCAATGAACCCCTCTGTTTTCTTGGAGGCTCTTGAA 1080
Db      361 LysLysAlaLysGluSerSerGluLysProPro-----GluLysProAlaLys 376
```



[illegible]

RESULT 10  
US-11-124-367A-442  
; Sequence 442, Application US/11124367A  
; Publication No. US20060024700A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: Hongjin Huang  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof  
; FILE REFERENCE: CU001519.ORD  
; CURRENT APPLICATION NUMBER: US/11/124,367A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,846  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/582,609  
; PRIOR FILING DATE: 2004-06-25  
; PRIOR APPLICATION NUMBER: US 60/599,554  
; PRIOR FILING DATE: 2004-08-09  
; NUMBER OF SEQ ID NOS: 34460  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 442  
; LENGTH: 2803  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-124-367A-442

Alignment Scores:  
Pred. No.: 0.0419 Length: 2803  
Score: 129.50 Matches: 71  
Percent Similarity: 42.3% Conservative: 72  
Best Local Similarity: 21.0% Mismatches: 119  
Query Match: 4.5% Indels: 76

```
US-11-124-367A-445
; Sequence 445, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; FILE REFERENCE: CLO01519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-08-09
; PRIOR APPLICATION NUMBER: US 60/599,554
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 445
; LENGTH: 2803
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-367A-445

Alignment Scores:
Pred. No.: 0.0419 Length: 2803
Score: 129.50 Matches: 71
Percent Similarity: 42.3% Conservative: 72
Best Local Similarity: 21.0% Mismatches: 119
Query Match: 4.5% Indels: 76
DB: 7 Gaps: 16

US-10-723-552-3_COPY_585_2156 (1-1572) x US-11-124-367A-445 (1-2803)
QY 712 CTCACGGGTTGGTGGAGTCCCTTATGTTCTCTCACACGAGCTACTTGGGATGTTGTCAA 771
Db 247 IleThrAlaLeuVal-----ValTrpLeuProAlaAsnProThr-----Glu 260
QY 772 CAGCTGTGCGACACTCTTCACACACGAGGAGCGAGAGAAATCTATTAGAGGCTAGA 831
Db 261 LysIleValArgValLeuPheProGlyAsnAlaProGlnAsnLysIleLeuGluGlyLeu 280
QY 832 AAAATGTTCTGGGCGGCAC-----GGGCGACCCACGCGGTTGCAAAATCAGATTAC 885
Db 281 GluLysLeuArgHisLeuAspPheLeuArgTyrProValAlaThrGlnLysAspLeuAla 300
QY 886 ATGGGATTTCCCTTAAC-----CGCCCGGTTGGGACTACACACGCGCTGAAGGTAGG 939
Db 301 SerGlyAlaValProThrAsnLeuLysProSerLysIleLysGlnArgAlaAspSerLys 320
QY 940 GAGAGCTTGAATAATCTATCGCAGGCTCTGGTGGCGGTCTC-----GTA 1020
Db 321 GluSerLeuLysAlaThrThrLysThrAlaValSerLysLeuAlaLysArgGluGluVal 340
QY 982 -----CGGGCGCTCAAGCGGCCCTACTAATTTGGCTAAG-----GTA 1020
Db 341 ValGluGluGlyAlaLysGluAlaArgSerGluLeuAlaLysGluLeuAlaLysThrGlu 360
QY 1021 AGAGAAGTGATGACGGGACCGAATCAACCCCTCTGTTTCTTGAGAGGCTCTTGGA 1080
Db 361 LysLysAlaLysGluSerSerGluLysProPro-----GluLysProAlaLys 376
QY 1081 GCCTTCAGCGGTACACCCCTTTTGATCCCACTCAGAGGCCCAAAAGCCCTCAGTGCT 1140
Db 377 ProGluArgValLysThrGluSerSerGluAlaLeuLysAlaGluLysArgLysLeuIle 396
QY 1141 TTGGCTTTATAGGACGACCTTGATATTAGAAAGAGCTTCAGAGCTCGAAGGG 1200
Db 397 LysAspLysValGlyLysHisLeu-----LysGluLysIleSerLysLeuGlu 414
QY 1201 TTACAGGAGGCTGAGTTACGTGATCTAGTGAAGGAGCGAGAAAGTATTACAAAAGG 1260
Db 1201 LysIleValArgValLeuPheProGlyAsnAlaProGlnAsnLysIleLeuGluGlyLeu 504

US-11-124-367A-443
; Sequence 443, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; FILE REFERENCE: CLO01519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 443
; LENGTH: 2984
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-367A-443

Alignment Scores:
Pred. No.: 0.0421 Length: 2984
Score: 129.50 Matches: 71
Percent Similarity: 42.3% Conservative: 72
Best Local Similarity: 21.0% Mismatches: 119
Query Match: 4.5% Indels: 76
DB: 7 Gaps: 16

US-10-723-552-3_COPY_585_2156 (1-1572) x US-11-124-367A-443 (1-2984)
QY 712 CTCACGGGTTGGTGGAGTCCCTTATGTTCTCTCACACGAGCTACTTGGGATGTTGTCAA 771
Db 471 IleThrAlaLeuVal-----ValTrpLeuProAlaAsnProThr-----Glu 484
QY 772 CAGCTGTGCGACACTCTTCACACACGAGGAGCGAGAGAAATCTATTAGAGGCTAGA 831
Db 485 LysIleValArgValLeuPheProGlyAsnAlaProGlnAsnLysIleLeuGluGlyLeu 504
```

```
QY 832 AAAAATGTTCTGGGGCCGAC-----GGGGACCCACCGCGTTCACAAATGAGATTGAC 885
   : : : : :
Db 505 GluLysLeuArgHisLeuAspPheLeuArgTyrProValAlaThrGlnLysAspLeuAla 524
   : : : : :
QY 886 ATGGGATTTCCTTAAC-----CGCCCGGTGGGACTACACACCGCTGAAGTAGG 939
   : : : : :
Db 525 SerGlyAlaValProThrAsnLeuLysProSerLysIleLysGlnArgAlaAspSerLys 544
   : : : : :
QY 940 GAGAGCTTGAATAATCTATCGCCAGGCTCTGGTGGCGGCTC----- 981
   : : : : :
Db 545 GluSerLeuLysAlaThrThyLysThrAlaValSerLysLeuAlaLysArgGluVal 564
   : : : : :
QY 982 -----CGGGCGCTCAAGACGGCCCACTAATTGGCTAAG-----GTA 1020
   : : : : :
Db 565 ValGluGluGlyAlaLysGluAlaArgSerGluLeuAlaLysGluLeuAlaLysThrGlu 584
   : : : : :
QY 1021 AGAAGATGATGACGGGACCAAGTAACCCCTCTGTTTCTTGAGAGCTCTTGAA 1080
   : : : : :
Db 585 LysLysAlaLysGluSerSerGluLysPro-----GluLysProAlaLys 600
   : : : : :
QY 1081 GCCTTCAGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAGCCTCAGTGCT 1140
   : : : : :
Db 601 ProGluArgValLysThrGluSerSerGluAlaLysAlaGluLysArgLysLeu 620
   : : : : :
QY 1141 TTGGCTTTATAGCACAGTCCCTCGATATAGAAAGAGCTTCAGAGCTCGAAGG 1200
   : : : : :
Db 621 LysAspLysValGlyLysHisLeu-----LysGluLysIleSerLysLeuGlu 638
   : : : : :
QY 1201 TTACAGGAGCTGATGATTACGTGTAGTGAAGGAGGACAGAAAGTATATTACAAAGG 1260
   : : : : :
Db 639 LysLysAspLysGluLysGluLysGluLysGluLysGluLys-----Lys 655
   : : : : :
QY 1261 GAGACAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1320
   : : : : :
Db 656 LysAspGluGlyArgLysGluLysGluLysAspAlaLysLysGluLys-----ArgLys 674
   : : : : :
QY 1321 CGTAATAACCGCAAGAAAGAAATTTGACTAAG----- 1353
   : : : : :
Db 675 AspThrLysProGluLeuLysLysIleSerLysProAspLeuLysProPheThrProGlu 694
   : : : : :
QY 1354 -----ATCTTGGCTGCGTGTGAAGGGAAAGCAATACGGAAGAGAGA 1401
   : : : : :
Db 695 ValArgLysThrLeuTyrLysAlaLysValProGlyArgValLysIleAspArgSerArg 714
   : : : : :
QY 1402 GATTTTAGG-----AAATAGGTGAGCCCTAGACAGCTCAGG 1440
   : : : : :
Db 715 AlaIleArgGlyGluLysGluLeuSerSerGluProGlnThrProProAlaGlnLysGly 734
   : : : : :
QY 1441 AACCTG-----GGCAATAG-----ACCCACTCGAC 1467
   : : : : :
Db 735 ThrValProLeuProThrIleSerGlyHisArgGluLeuValLeuSerSerProGluAsp 754
   : : : : :
QY 1468 -----AAGACCAATGTGCATATTGTAAGAAAGA 1497
   : : : : :
Db 755 LeuThrGlnAspPheGluGluMetLysArgGluGluArgAlaLeuLeuAlaGlnArg 774
   : : : : :
QY 1498 GGACACTGGGCAAGCAAGTCCCA-----AGAAGGGAACCAAGACCAAGA 1546
   : : : : :
Db 775 -AspThrGlyLeuGlyAspLysProPheProLeuAspThrAlaGluGly 791
   : : : : :
```

## RESULT 13

```
US-11-124-367A-441
; Sequence 441, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
```

```
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 441
; LENGTH: 3027
; TYPE: PRF
; ORGANISM: Homo sapiens
US-11-124-367A-441
```

## Alignment Scores:

Pred. No.:	0.0421	Length:	3027
Score:	129.50	Matches:	71
Percent Similarity:	42.3%	Conservative:	72
Best Local Similarity:	21.0%	Mismatches:	119
Query Match:	4.5%	Indels:	76
DB:	7	Gaps:	16

US-10-723-552-3\_COPY\_585\_2156 (1-1572) x US-11-124-367A-441 (1-3027)

```
QY 712 CTCACGGGTGGTGGAGTCCCTTATGTTCTCTCACGAGCTACTTGGGATGTGTCAA 771
   : : : : :
Db 471 IleThrAlaLeuVal-----ValTripleuProAlaAsnProThr-----Glu 484
   : : : : :
QY 772 CAGCTGTGCAGACACTTTCACAAACGAGGAGCGAGAGAGAAATTCATTAGAGCTAGA 831
   : : : : :
Db 485 LysIleValArgValLeuPheProGlyAsnAlaProGlnAsnLysIleLeuGluGlyLeu 504
   : : : : :
QY 832 AAAAATGTTCTCTGGGGCCGAC-----GGGGACCCACCGCGTTCACAAATGAGATTGAC 885
   : : : : :
Db 505 GluLysLeuArgHisLeuAspPheLeuArgTyrProValAlaThrGlnLysAspLeuAla 524
   : : : : :
QY 886 ATGGGATTTCCCTTAAC-----CGCCCGGTGGGACTACACACCGCTGAAGTAGG 939
   : : : : :
Db 525 SerGlyAlaValProThrAsnLeuLysProSerLysIleLysGlnArgAlaAspSerLys 544
   : : : : :
QY 940 GAGAGCTTGAATAATCTATCGCCAGGCTCTGGTGGCGGCTC----- 981
   : : : : :
Db 545 GluSerLeuLysAlaThrThyLysThrAlaValSerLysLeuAlaLysArgGluVal 564
   : : : : :
QY 982 -----CGGGCGCTCAAGACGGCCCACTAATTGGCTAAG-----GTA 1020
   : : : : :
Db 565 ValGluGluGlyAlaLysGluAlaArgSerGluLeuAlaLysGluLeuAlaLysThrGlu 584
   : : : : :
QY 1021 AGAAGATGATGACGGGACCGAATGAAACCCCTCTCTTTTCTTGAGAGCTCTTGAA 1080
   : : : : :
Db 585 LysLysAlaLysGluSerSerGluLysPro-----GluLysProAlaLys 600
   : : : : :
QY 1081 GCCTTCAGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAGCCTCAGTGCT 1140
   : : : : :
Db 601 ProGluArgValLysThrGluSerSerGluAlaLysAlaGluLysArgLysLeu 620
   : : : : :
QY 1141 TTGGCTTTATAGCACAGTCCCTCGATATAGAAAGAGCTTCAGAGCTCGAAGG 1200
   : : : : :
Db 621 LysAspLysValGlyLysHisLeu-----LysGluLysIleSerLysLeuGlu 638
   : : : : :
QY 1201 TTACAGGAGCTGATGATTACGTGTAGTGAAGGAGGACAGAAAGTATATTACAAAGG 1260
   : : : : :
Db 639 LysLysAspLysGluLysGluLysGluLysGluLysGluLys-----Lys 655
   : : : : :
QY 1261 GAGACAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1320
   : : : : :
Db 656 LysAspGluGlyArgLysGluLysGluLysAspAlaLysLysGluLys-----ArgLys 674
   : : : : :
QY 1321 CGTAATAACCGCAAGAAAGAAATTTGACTAAG----- 1353
   : : : : :
Db 675 AspThrLysProGluLeuLysLysIleSerLysProAspLeuLysProPheThrProGlu 694
   : : : : :
QY 1354 -----ATCTTGGCTGCGTGTGAAGGGAAAGCAATACGGAAGAGAGA 1401
   : : : : :
Db 695 ValArgLysThrLeuTyrLysAlaLysValProGlyArgValLysIleAspArgSerArg 714
   : : : : :
QY 1402 GATTTTAGG-----AAATAGGTGAGCCCTAGACAGCTCAGG 1440
   : : : : :
Db 715 AlaIleArgGlyGluLysGluLeuSerSerGluProGlnThrProProAlaGlnLysGly 734
   : : : : :
QY 1441 AACCTG-----GGCAATAG-----ACCCACTCGAC 1467
   : : : : :
Db 735 ThrValProLeuProThrIleSerGlyHisArgGluLeuValLeuSerSerProGluAsp 754
   : : : : :
QY 1468 -----AAGACCAATGTGCATATTGTAAGAAAGA 1497
   : : : : :
Db 755 LeuThrGlnAspPheGluGluMetLysArgGluGluArgAlaLeuLeuAlaGlnArg 774
   : : : : :
QY 1498 GGACACTGGGCAAGCAAGTCCCA-----AGAAGGGAACCAAGACCAAGA 1546
   : : : : :
Db 775 -AspThrGlyLeuGlyAspLysProPheProLeuAspThrAlaGluGly 791
   : : : : :
```

```
Db      695 ValArgLysThrLeuTyrLysAlaLysValProGlyArgValLysIleAspArgSerArg 714
      :::      |||      |||      |||      |||      |||      |||      |||
Qy      1402 GATTTAGG-----AAATTAGGTGAGGCGCCCTAGACAGTCAGG 1440
      |||      |||      |||      |||      |||      |||      |||
Db      715 AlaIleArgGlyGluLysGluLysSerGluProGlnThrProProAlaGlnLysGly 734
      |||      |||      |||      |||      |||      |||      |||
Qy      1441 AACCTG-----GGCAATAG-----ACCCACTCGAC 1467
      |||      |||      |||      |||      |||      |||      |||
Db      735 ThrValProLeuProThrIleSerGlyHisArgGluLeuValLeuSerSerProGluAsp 754
      |||      |||      |||      |||      |||      |||      |||
Qy      1468 -----AAGGACCAATGTCATATTGTAAGAAAGA 1497
      |||      |||      |||      |||      |||      |||      |||
Db      755 LeuThrGlnAspPheGluGluMetLysArgGluGluArgAlaLeuLeuAlaGluGlnArg 774
      |||      |||      |||      |||      |||      |||      |||
Qy      1498 GGACACTGGCAAGGAATGCCCA---AGAAAGGAAACAAAGGACCAAGGA 1546
      |||      |||      |||      |||      |||      |||      |||
Db      775 -AspThrGlyLeuGlyAspLysProPheProLeuAspThrAlaGluGluGly 791
      |||      |||      |||      |||      |||      |||      |||

RESULT 14
US-11-182-016-22
; Sequence 22, Application US/11182016
; Publication No. US20060019294A1
; GENERAL INFORMATION:
; APPLICANT: SUGEN, INC.
; TITLE OF INVENTION: TYROSINE KINASE SUBSTRATE (TKS) PROTEINS
; FILE REFERENCE: 038602/0102
; CURRENT APPLICATION NUMBER: US/11/182,016
; CURRENT FILING DATE: 2005-07-15
; PRIOR APPLICATION NUMBER: US/09/958,359
; PRIOR FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 1744
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Grub
US-11-182-016-22

Alignment Scores:
Pred. No.: 0.0442 Length: 1744
Score: 129.00 Matches: 124
Percent Similarity: 33.0% Conservative: 63
Best Local Similarity: 21.9% Mismatches: 162
Query Match: 4.5% Indels: 217
DB: 7 Gaps: 30

US-10-723-552-3_copy_585_2156 (1-1572) x US-11-182-016-22 (1-1744)
Qy      213 TGGACCCGGCTCTCATCCCGATCAGGAGCCCTATATCTTACGTG---GCAAGATTGGC 269
      |||      |||      |||      |||      |||      |||      |||
Db      670 TrpSerGlnLeuArg-----AlaLeuPheLeuIleValLeuSerThrTrpGly 685
      |||      |||      |||      |||      |||      |||      |||
Qy      270 AGAGATCTCGCCATGGTTAAACCATGGCTGTAATAGCCAAAGAACCCAGGTCCTCCG 329
      |||      |||      |||      |||      |||      |||      |||
Db      686 GlnGlyAlaMetGly-TrpGlyTrpProTrp-----ProAr 697
      |||      |||      |||      |||      |||      |||      |||
Qy      330 AATTCTGGCTCTTGAGAGAAACAAACACTCGGCTGAAAAGTCAAGCCCTCTCTCA 389
      |||      |||      |||      |||      |||      |||      |||
Db      697 9-----AlaAlaGlyGluGlnGlnGlyAlaGlyArgGlyAlaAlaGlyProPr 715
      |||      |||      |||      |||      |||      |||      |||
Qy      390 TATCTACCCCGAGATTGAGGAG-----CCACCGGCTTGGCCGGAACCCCAATCT-- 438
      |||      |||      |||      |||      |||      |||      |||
Db      715 oLeuGlyProSerThrAlaGlyThrValCysProProSerSerValAlaProTrpAlaAr 735
      |||      |||      |||      |||      |||      |||      |||
Qy      439 -----GTTCCTCCACCCCTTATCTGGCACAGGCTGCGCGAGGGG 479
      |||      |||      |||      |||      |||      |||      |||
Db      735 gLeuLeuSerGlnValLeuLeuAspProProLeuLeuLeuAlaAlaGlyLeuAlaAla 755
      |||      |||      |||      |||      |||      |||      |||
Qy      480 ACCCTTTGCCCT-----CTGGAGCTCCGCGGTGGA 512
      |||      |||      |||      |||      |||      |||      |||
```

```
Db      755 aproPheSerGlnGlyCysTrpAlaAlaGlnValGlnGlyGlnGlyProHisLeuLeuG1 775
Qy      513 GGGACCTGCTCGAGGACTCGAGCCGAGGGGGCCACCCCGGAGCGGACAGACAGAT 572
      |||      |||      |||      |||      |||      |||      |||
Db      775 uGly-----AlaGlyThr-GlyLeuTrpGlyProPro-----GlnGlyGln----- 788
      |||      |||      |||      |||      |||      |||      |||
Qy      573 CGCGACATTACCGCTGGCGACGTACGGCCCTCCACACCGGGGGCCCAATTGCGACCCCT 632
      |||      |||      |||      |||      |||      |||      |||
Db      789 -----CysSerGlnHisSerLeuProThrArgProSerSerCysGlnProS 804
      |||      |||      |||      |||      |||      |||      |||
Qy      633 CCAAGT-----ATTGGCCCTTTCTTCGACAGATCTCTATAATTGGAATAACCATCC 686
      |||      |||      |||      |||      |||      |||      |||
Db      804 erLysAlaLeuIleHisProPheMet-Cys----- 813
      |||      |||      |||      |||      |||      |||      |||
Qy      687 CCTTTCTCGAGAGATCCCAACCGCTCACGGGGTGGTGAGTCCCTTATGTCTCTCA 746
      |||      |||      |||      |||      |||      |||      |||
Db      814 -----LeuLeuGluGluAlaLeuProGlnGlyPro-MetGluAlaThrLeuAsnLeuLe 831
      |||      |||      |||      |||      |||      |||      |||
Qy      747 CCAGCTACTTGGGATGATTGTCACACAGCTCTCTG----- 780
      |||      |||      |||      |||      |||      |||      |||
Db      831 uGlnProLeuPheLysAspIleGlnHisLeuLeuGlyGlyGlyIleLeuLeuGlnSerG1 851
      |||      |||      |||      |||      |||      |||      |||
Qy      781 -----CACACACTCTTC-----ACAACCGAGGACGAGAGA 809
      |||      |||      |||      |||      |||      |||      |||
Db      851 nGlyLeuProCysLeuGlyGlnAlaLeuLeuAspLeuSerAlaGluSerThrGluSerG1 871
      |||      |||      |||      |||      |||      |||      |||
Qy      810 GAGATTCTATTAGAGCTAGAAAAAATGTTCTCGGGCCGACCGGCGCCGCGTT 869
      |||      |||      |||      |||      |||      |||      |||
Db      871 nLeuCysLeuLeuGlnGlyArgGlnGlyValPro-----ArgHis-SerLysAlaC 888
      |||      |||      |||      |||      |||      |||      |||
Qy      870 GCAAAATGAGATTGACATGG-----CATTTCCCTCTTAAC 902
      |||      |||      |||      |||      |||      |||      |||
Db      888 ysGlnLeuLeuProTrpAlaArgGluProLeuGlnHisLeuLeuGluSerProGlnA 908
      |||      |||      |||      |||      |||      |||      |||
Qy      903 TCGCCCGGTTGGAGCTACAAACACGGCTGAAGGTAGGAGAGCTTGAATAATCTATCGCCA 962
      |||      |||      |||      |||      |||      |||      |||
Db      908 laLeuProLeuLeuLeuLeuHisValGlnValGlyGluAlaHisLysLeuValAsnG 928
      |||      |||      |||      |||      |||      |||      |||
Qy      963 GGCTC-----TGGTGGCGGCTC-----T 980
      |||      |||      |||      |||      |||      |||      |||
Db      928 lyLeuValHisPheLeuIleGlnCysSerTrpAlaLeuAlaLeuGlnLeuAlaAlaGlys 948
      |||      |||      |||      |||      |||      |||      |||
Qy      981 CCGGGCGCCT-----CAAGACGCGCCACTAATTTGGCTAAGGTAAAGAGATGATGCA 1034
      |||      |||      |||      |||      |||      |||      |||
Db      948 erGlyAlaGlnProHisGlnAspGlyProProlleProLeuGln-----CysA 964
      |||      |||      |||      |||      |||      |||      |||
Qy      1035 GGGACCGGAATGAACCCCTCTGTTTCTTGAAGAGGCTCTTGAAGGCTTCAGCGGTA 1094
      |||      |||      |||      |||      |||      |||      |||
Db      964 rg-----GlnProGlyIleCysG 970
      |||      |||      |||      |||      |||      |||      |||
Qy      1095 CACCCCTTTTGATCCCACTCAGAGGCCCAAAAGCCTCAGTGG----- 1138
      |||      |||      |||      |||      |||      |||      |||
Db      970 lnHisLeuLeu-----GlnTrpLeuGlyHisSerHisC 981
      |||      |||      |||      |||      |||      |||      |||
Qy      1139 -----CTTTGG-----CCTTTATAGCACAGTCAGCCTCGATAT 1172
      |||      |||      |||      |||      |||      |||      |||
Db      981 ysLeuLeuPheLeuTrpLeuCysCysSerLeuGlnLeuAsnGlyLeuProTrpArgA 1001
      |||      |||      |||      |||      |||      |||      |||
Qy      1173 TAGAAGAAGCTTCAGAGACTGGAAGGTTTACAGAGGCTGAGTTACTGATCTAGTGA 1232
      |||      |||      |||      |||      |||      |||      |||
Db      1001 laGlnHisThrLeuAlaGlnAlaThrProTrpHisPheLeuValTyrLeu----- 1017
      |||      |||      |||      |||      |||      |||      |||
Qy      1233 GGAGGCAGAAAGTATATTACAAAGGGAGA----- 1264
      |||      |||      |||      |||      |||      |||      |||
Db      1018 -----TyrProValThrGlyArgGlyValProValThrSerLeuProLeuG 1033
      |||      |||      |||      |||      |||      |||      |||
Qy      1264 ----- 1264
      |||      |||      |||      |||      |||      |||      |||
Db      1033 lnLeuLeuTrpLeuGlyHisSerPheGlnIlelleLeuGlnHisLeuSerThrLeuLys 1053
      |||      |||      |||      |||      |||      |||      |||
Qy      1265 -----CAGAAGAAAGAAAGAAACAAAGAAAGAGAGAGAGAGAGAGAGGA 1313
      |||      |||      |||      |||      |||      |||      |||
Db      1053 erThrGluPheSerTrpLysValIleMetAsnGluAsnGlnGlnProLeuAsnLysGlyA 1073
      |||      |||      |||      |||      |||      |||      |||
```

```
Qy 1314 GGAAGACGTAAATA-----AACGCAAGAGAGAATTGACTAAGATCTTGGCTGC 1364
Db 1073 rgileSerValLeuLysLeuAlaGlnGlyArgAsn-----GluCysTrpArgG 1089
Qy 1365 AGTCG-----TTGAAGGGAA 1379
Db 1089 InTrpArgCysLeuThrLysValGlnGlnAspGlyGlnProGlnCysLeuIleArgProG 1109
Qy 1380 AAGCAATACGAAAGAGAGAGAGATTTAGGAAATTA-----GGTCAGGCCCTAGACA 1433
Db 1109 luGlnValGluValMetLysSerCysValGlnArgValSerGlyGlyTrpGlyLeuLeuA 1129
Qy 1434 GTCAGGGAACCTGGCAATAGGACCCCACTGCAAGAGCAATGTGCATATTGTAAGA 1493
Db 1129 rgArgAlaArgTrpGlyAsnGlyGlnGlnSerSerProThrLeuValHisProSerArg 1149
Qy 1494 AAGAGGAC 1501
Db 1149 InAspGlu 1151

RESULT 15
US-10-982-545-14
; Sequence 14, Application US/10982545
; Publication No. US20050244890A1
; GENERAL INFORMATION:
; APPLICANT: Davies, Huw Alun
; APPLICANT: McGuire, James
; APPLICANT: Simonsen, Anja Hviid
; APPLICANT: Blennow, Kaj
; APPLICANT: Podust, Vladimir
; APPLICANT: CIPHERGEN Biosystems, Inc.
; TITLE OF INVENTION: Biomarkers for Alzheimer's Disease
; FILE REFERENCE: 016866-011550US
; CURRENT APPLICATION NUMBER: US/10/982,545
; CURRENT FILING DATE: 2004-11-06
; PRIOR APPLICATION NUMBER: US 60/518,360
; PRIOR FILING DATE: 2003-11-07
; PRIOR APPLICATION NUMBER: US 60/526,753
; PRIOR FILING DATE: 2003-12-02
; PRIOR APPLICATION NUMBER: US 60/546,423
; PRIOR FILING DATE: 2004-02-19
; PRIOR APPLICATION NUMBER: US 60/547,250
; PRIOR FILING DATE: 2004-02-23
; PRIOR APPLICATION NUMBER: US 60/558,896
; PRIOR FILING DATE: 2004-04-02
; PRIOR APPLICATION NUMBER: US 60/572,617
; PRIOR FILING DATE: 2004-05-18
; PRIOR APPLICATION NUMBER: US 60/586,503
; PRIOR FILING DATE: 2004-07-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: nerve growth factor inducible Neurosecretory
; OTHER INFORMATION: protein vaccinia virus growth factor (VGF)
; OTHER INFORMATION: precursor
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(22)
; OTHER INFORMATION: signal peptide
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (23)..(615)
; OTHER INFORMATION: nerve growth factor inducible Neurosecretory
; OTHER INFORMATION: protein vaccinia virus growth factor (VGF)
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (373)..(417)
; OTHER INFORMATION: biomarker peptide 4808 Da, 4812 Da (Q10),
```

```
; OTHER INFORMATION: processed fragment of nerve growth factor
; OTHER INFORMATION: inducible Neurosecretory protein vaccinia virus
; OTHER INFORMATION: growth factor (VGF)
US-10-982-545-14
```

```
Alignment Scores:
Pred. No.: 0.045 Length: 615
Score: 128.50 Matches: 100
Percent Similarity: 32.8% Conservative: 44
Best Local Similarity: 22.8% Mismatches: 176
Query Match: 4.5% Indels: 119
DB: 6 Gaps: 19
```

```
US-10-723-552-3_COPY_585_2156 (1-1572) x US-10-982-545-14 (1-615)
Qy 412 CCACCGCTGGCGGGAACCCCAATCTGTCCCCACCCCTTATCTGGCACAGGGTGC 471
Db 24 ProProGlyArgProGluAlaGln-----ProProLeuSerSerGluHisLys 40
Qy 472 GCGAGGGGACCTTTTGCCCTCTCGGAGCTCCGCGGTGGAGGACCTGTGCAGGGACT 531
Db 41 -----GluProValAlaGlyAspAlaValProGlyProLysAspGlySer 55
Qy 532 -----CGAGCGCGAGGGCGCCACCCCGGAGCGGACAGACGAG----- 570
Db 56 AlaProGluValArgGlyAlaArgAsnSerGluProGlnAspGluGlyGluLeuPheGln 75
Qy 571 -----ATCGCGACATTACCCTG-----CCACGCTACGGC 600
Db 76 GlyValAspProArgAlaLeuAlaValLeuLeuGlnAlaLeuAspArgProAlaSer 95
Qy 601 CTTCCACACCGGGGGCCCAATTGCAG----- 627
Db 96 ProProAlaProSerGlySerGlnGlnGlyProGluGluGluAlaGluAlaLeuLeu 115
Qy 628 -----CCCTCCAGTATTGGCCCTTTCTCTCTCCAGGAGTATCTCT 654
Db 116 ThrGluThrValArgSerGlnThrHisSerLeuProAlaProGluSerProGluProAla 135
Qy 655 GCAGATCTCTATAATTGGAAACTAACCATCCCTCTCTCGGAGGATCCC----- 705
Db 136 AlaProArgProGlnThrProGluAsnGlyProGluAlaSerAspProSerGluGlu 155
Qy 706 ---CAACGCTCAGCGGTGGTGAGTCCCTTATGTCTCTCACCAGCCCTACTTGGAT 762
Db 156 LeuGluAlaLeuAlaSerLeuLeuGlnGluLeuArg---AspPheSerProSerSerAla 174
Qy 763 GATTGTCAACAGCTGCTGCAGACACTCTTTCACACCGAGGAGCGAGAGAAATTCATT 822
Db 175 LysArgGlnGlnGluThrAlaAlaAla---GluThrGluThrArgThrHisThrLeuThr 193
Qy 823 GAGGCTAGAAAAAATCTTCTGGGGCGGAC----- 852
Db 194 ArgValAsnLeuGluSerProGlyProGluArgValTrpArgAlaSerTrpGlyGluPhe 213
Qy 853 -----GGCGCACCACCGGTTGCAAAAT 876
Db 214 GlnAlaArgValProGluArgAlaProLeuProProProAlaProSerGlnPheGlnAla 233
Qy 877 GAGATTGACATGGGATTTCCCTTAACCTCGCCCGGTGGGACTTACACACGGCTGAAGGT 936
Db 234 ArgMetProAspSerGlyProLeuProGlu-----ThrHisLysPheGlyGluGly 250
Qy 937 AGGAGAGCTTGAANAATCTATCGCCAGGCTCTGTGCGGGTCTCCGGGCGGCC----- 990
Db 251 ValSerSerProLysThrHisLeuGlyGluAlaLeuAlaProLeuSerLysAlaTyrGln 270
Qy 991 -----TCAAGACGCGCCCACTAAATTTGGCTAAGGTAAGAGAA 1026
Db 271 GlyValAlaAlaProPheProLysAlaArgArgPro-----GluSer 284
Qy 1027 GTGATCAGGGACCGAATGAACCCCTCTGTGTTTTTCTTGAGAGGCTCTTG----- 1077
Db 1027 GTGATCAGGGACCGAATGAACCCCTCTGTGTTTTTCTTGAGAGGCTCTTG----- 1077
```

```
Db      285 AlaLeuLeuGlyGlySerGluAlaGlyGluArgLeuLeuGlnGlnGlyLeuAlaGlnVal 304
Qy      1078 GAAGCCTTCAGCGGTACACCCCTTTTGATCCACCTCAGAGGCCCAAAAAGCCTCAGTG 1137
Db      305 GluAlaGlyArgArgGlnAlaGluAlaThrArgGlnAlaAlaAlaGlnGluGluArgLeu 324
Qy      1138 GCT---TTGGCCTTTATAGGACAGTCAGCCTTTGGATATTAGAAAGAGCTTCAG----- 1188
Db      325 AlaAspLeuAla-----SerAspLeuLeuLeuGlnTyrLeuLeuGlnGlyGly 340
Qy      1189 ---AGACTGGAGGTTTACAGGAGGCTGAGTTACGTCTAGTCTAGTCAAGGAGGCAGAGAAA 1245
Db      341 AlaArgGlnArgGlyLeuGlyArgGlyLeuGlnGluAlaAlaGluGluArgGluSer 360
Qy      1246 GTATATTACAAGAGGAGACAGAGAGAAAGGGAACAAAGAAAGAGAGAGAAAGAGAG 1305
Db      361 AlaArgGluGluGluGluAlaGluGlnGluArgArgGlyGlyGluGluArgValGlyGlu 380
Qy      1306 GAAAGGGAGGAAAGACGTTAATAACGGCAAGAG-----AAGAAT 1344
Db      381 GluAspGluGluAlaAlaGluAlaGluAlaGluAlaGluAlaGluArgAlaArgGln 400
Qy      1345 TTGACTAAGATCTTGGCTGCAGTGGTTGAAGGGAAAAGCAATACGGAAGAGAGAGAGAT 1404
Db      401 AsnAlaLeuLeuPheAlaGluGluGluAspGlyGluAlaGlyAlaGluAspLysArgSer 420
Qy      1405 -----TTTAGGAAAATTAGTCAAGCCCTAGACAGCTCAGGG 1440
Db      421 GlnGluGluThrProGlyHisArgArgLysGluAlaGluGlyThrGluGluGlyGly 439
```

Search completed: February 14, 2006, 17:20:48  
Job time : 48.8924 secs

GenCore version 5.1.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 14:51:21 ; Search time 20.0582 Seconds  
(without alignments)  
3295.451 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_2307\_5741  
Perfect score: 6183  
Sequence: 1 ATGGTGCCACAGGCAACA.....CTGTCAATAACCTCTCAGAC 3435

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB spool/US10723552/runat 14022006 125143 12937/app query.fasta\_1  
-DB-PIR -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USBR=US10723552 @CGN 1.1 107 @runat 14022006 125143 12937 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 80:\*  
1: Pirl1:\*  
2: Pirl2:\*  
3: Pirl3:\*  
4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4135.5	66.9	1165	1	GNLJGL HIV-1 retropepsin
2	3836	62.0	1196	1	GNMVG HIV-1 retropepsin
3	3818.5	61.8	1204	2	S35475 pol polyprotein -
4	3809.5	61.6	1204	2	S70393 pol polyprotein -
5	3802.5	61.5	1189	1	GNMW7 HIV-1 retropepsin
6	3797.5	61.4	1199	1	GNWIM HIV-1 retropepsin
7	3778	61.1	1196	1	GNMVR HIV-1 retropepsin
8	3696	59.8	1784	2	T10532 gag-pol polyprotei
9	3536.5	57.2	1046	1	GNMVE pol polyprotein -
10	2639.5	42.7	843	1	GNVWK pol polyprotein -
11	1725.5	27.9	476	2	S04842 pol polyprotein -
12	1681	27.2	581	2	A42743 pol polyprotein -
13	1572	25.4	1751	2	T09394 gag-pro-pol polypr
14	1147	18.6	409	2	S12588 pol polyprotein -

15	1142.5	18.5	602	2	B46312 pol polyprotein -
16	1094	17.7	559	2	A46311 pol polyprotein -
17	1084	17.5	390	2	A26621 retrovirus-related
18	1029.5	16.7	294	2	A05071 pol polyprotein -
19	1023.5	16.6	848	4	A44282 retrovirus-related
20	984.5	15.6	282	2	A26103 pol polyprotein -
21	905.5	14.6	473	2	A03959 pol polyprotein -
22	811	13.1	1157	1	GNLJLK pol polyprotein -
23	801	13.0	1161	2	S18738 pol protein - simi
24	692	11.2	886	1	GNLJSP pol polyprotein -
25	678.5	11.0	958	2	S15566 pol protein - simi
26	662.5	10.7	1182	2	T29097 pro-pol-dutPase po
27	639	10.3	982	1	GNLJH2 pol polyprotein -
28	604	9.8	1053	1	GNLJBT HIV-1 retropepsin
29	601	9.7	1146	1	GNLJ22 HIV-1 retropepsin
30	601	9.7	1145	1	GNLJEM HIV-1 retropepsin
31	599.5	9.7	1145	1	GNLJEV HIV-1 retropepsin
32	587	9.5	867	1	GNLJMP pol polyprotein (c
33	578	9.3	867	1	GNLJSA pol polyprotein -
34	573.5	9.3	896	1	GNLJGH pol polyprotein -
35	570	9.2	697	2	A26132 gag-abl-pol polypr
36	567.5	9.2	896	1	GNLJCN pol polyprotein ho
37	563.5	9.1	1295	2	S60179 pol polyprotein ho
38	561.5	9.1	1611	2	G84493 probable retroelem
39	561.5	9.1	2272	2	T18572 gag, pol and env p
40	552.5	8.9	399	4	A45689 retrovirus-related
41	543.5	8.8	852	1	GNLJGA pol polyprotein -
42	541.5	8.8	852	2	S29358 pol protein - bovi
43	538.5	8.7	888	1	GNLJHD pol polyprotein -
44	538	8.7	843	2	S33123 pol polyprotein -
45	531.5	8.6	1313	2	T29193 hypothetical prote

ALIGNMENTS

RESULT 1

GNLJGL

HIV-1 retropepsin (EC 3.4.23.16) - gibbon ape leukemia virus

N;Contains: nuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA polymer

C;Species: gibbon ape leukemia virus

C;Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 09-Jul-2004

C;Accession: B32595

R;Delassus, S.; Sonigo, P.; Wain-Hobson, S.

Virology 173, 205-213, 1989

A;Title: Genetic organization of gibbon ape leukemia virus.

A;Reference number: A32595; MUID:90051069; PMID:2683360

A;Accession: B32595

A;Molecule type: genomic RNA

A;Residues: 1-1165 <DEL>

A;Cross-references: UNIPROT:P21414; UNIPARC:UPI0000131EDE; GB:M26927; NID:g332610; PIDN

C;Comment: The pol polyprotein contains reverse transcriptase and possibly an endonucle

C;Genetics:

A;Gene: pol

C;Superfamily: pol polyprotein

C;Keywords: aspartic proteinase; endonuclease; hydrolase; nucleotidyltransferase; polyp

F;3-102/Product: retropepsin #status predicted <RTP>

F;27/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:

Pred. No.:	1.12e-268	Length:	1165
Score:	4135.50	Matches:	777
Percent Similarity:	79.7%	Conservative:	128
Best Local Similarity:	68.4%	Mismatches:	200
Query Match:	66.9%	Indels:	31
DB:	1	Gaps:	7

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x GNLJGL (1-1165)

Qy 4 GGTGCCACAGGCAACAACAGTATCCATGGACTACCGAAGAACAGTTGACTTGGGAGTG 63

Db 51 GlyAlaThrGlySerValTyPrThrLysArgLeuLeuLysIleGlyHis 70

Qy 64 GGACGGGTAAACCCACTCGTTTCTGGTCATACCTGAGTGAGTCCGCCAGCACCCCTCTTAGGTAGA 123

Db 71 LysGlnValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGlyArg 90  
:::|||||  
QY 124 GACTTATGACCAAGATGGGACCAAAATTTCTTTTGNACAGGGAACCAAGAGTGTCT 183  
|||||  
Db 91 AspLeuLeuThrLysLeuAlaGlnIleGlnPheSerAlaGluGlyProGlnValThr 110  
|||||  
QY 184 GCAAATAACAACCTATCACTGTGTGACCTCCAAATTTAGATGACGAATATCGACTATAC 243  
|||||  
Db 111 TrpGlyGluArgProThrMetCysLeuValLeuAsnLeuGluGluTyArgLeuHis 130  
|||||  
QY 244 TCTCCCTAGTAAAGCCTGTATCAAAATATACAATTTCTGGTTGGAACAGATTTCCCAAGCC 303  
|||||  
Db 131 GluLysProVal---ProSerSerIleAspProSerTrpLeuLeuPheProThrVal 149  
|||||  
QY 304 TGGGAGAAACCGAGGATGGTTTGGCAAGCAAGTTCCCCCAAGTATTTCACACTG 363  
|||||  
Db 150 TrpAlaGluArgAlaGlyMetGlyLeuAlaAsnGlnValProProValValValGluLeu 169  
|||||  
QY 364 AAGGCCAGTGCACACAGTGTCACTCAGACAGTACCCCTTGAGTAAAGAGAGCTCAAGAA 423  
|||||  
Db 170 ArgSerGlyAlaSerProValAlaValArgGlnTyPrometSerLysGluAlaArgGlu 189  
|||||  
QY 424 GGAATTCGGCGCATGTCCAAAGATTAATCCAACAGGCGCATCTAGTTCTCTGTCCTCAATCT 483  
|||||  
Db 190 GlyIleArgProHisIleGlnLysPheLeuAspLeuGlyValLeuValProCysArgSer 209  
|||||  
QY 484 CCCTGGAATATCTCCCTGCTACCGGTTAGAAAGCCTGGAGCTATGACTATCGACAGTA 543  
|||||  
Db 210 ProTrpAsnThrProLeuLeuProValLysProGlyThrAsnAspTyArgProVal 229  
|||||  
QY 544 CAGAGCTTGAGAGGTCAATAAACGGGTGAGGATATACACCAACAGTCCCGAACCT 603  
|||||  
Db 230 GlnAspLeuArgGluIleAsnLysArgValGlnAspIleHisProThrValProAsnPro 249  
|||||  
QY 604 TATAACCTCTGTGTGCTCTCCACCCCAACGGAGCTGGTATACAGTATTGGACTTAAAG 663  
|||||  
Db 250 TyrAsnLeuLeuSerSerLeuProProSerTyThrTrpTySerValLeuAspLeuLys 269  
|||||  
QY 664 GATGCTTCTTCTGCTGAGATTACACCCCTAGTACCAACCACTTTTGGCTTGAATGG 723  
|||||  
Db 270 AspAlaPhePheCysLeuArgLeuHisProAsnSerGlnProLeuPheAlaPheGluTrp 289  
|||||  
QY 724 AGAGATCCAGGTACCGGAGAACCGGCGAGCTCACCTGACCCGACTCCCAAGGGTTC 783  
|||||  
Db 290 LysAspProGluLysGlyAsnThrGlyGlnLeuThrTrpThrArgLeuProGlnGlyPhe 309  
|||||  
QY 784 AAGAACTCCCGACCATCTTTGACGAAGCCCTTACACAGAGACCTGGCCCAACTTCAGGATC 843  
|||||  
Db 310 LysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaProPheArgAla 329  
|||||  
QY 844 CAACACCTCAGGTGACCTCTCCAGTACGTGGATGACCTGCTCTCGGGGAGCCACC 903  
|||||  
Db 330 LeuAsnProGlnValValLeuLeuGlnTyValAspAspLeuLeuValAlaAlaProThr 349  
|||||  
QY 904 AAACAGGACTGTTAGAGGACGAGGACCTACTGCTGGAATCTCTGACCTAGGCTAC 963  
|||||  
Db 350 TyrGluAspCysLysGlyThrGlnLysLeuLeuGlnLeuSerLysLeuGlyTy 369  
|||||  
QY 964 AGAGCTCTGCTGAAGGCCAGATTTGACGAGAGAGGTAACTACTTGGGGTACAGT 1023  
|||||  
Db 370 ArgValSerAlaLysLysAlaGlnLeuCysGlnArgGluValThrTyLeuGlyTyLeu 389  
|||||  
QY 1024 TTGCGGAGCGGACGATGGCTGACGAGGACCGAGGACGAAACCTGTAGTCCAGATACG 1083  
|||||  
Db 390 LeuLysGluGlyLysArgTrpLeuThrProAlaArgLysAlaThrValMetLysIlePro 409  
|||||  
QY 1084 GCCCAACACAGCCAAACAAATGACAGATTTTGGGACAGCTGGATTTTGCAGACTG 1143  
|||||  
Db 410 ValProThrThrProArgGlnValArgGluPheLeuGlyThrAlaGlyPheCysArgLeu 429  
|||||  
QY 1144 TGGATCCCGGGGTTTGGACCTTAGCAGCCCACTCTACCCCTCAACCAAGAAAAAGG 1203  
|||||

Db 430 TrpIleProGlyPheAlaSerLeuAlaAlaProLeuTyProLeuThrLysGluSerIle 449  
|||||  
QY 1204 GAATTCCTCCCTGGCTCTCTGAGCACCAGAGCGCATTTGATGCTATCAAAAAGCCCTGCTG 1263  
|||||  
Db 450 ProPheIleTrpThrGluGluHisGlnAlaPheAspHisIleLysLysAlaLeuLeu 469  
|||||  
QY 1264 AGCGCACCTCTCTGGCCCTCCCTGACGTAACTAAACCCCTTTACCTTTATGTGGATGAG 1323  
|||||  
Db 470 SerAlaProAlaLeuAlaLeuProAspLeuThrLysProPheThrLeuTyIleAspGlu 489  
|||||  
QY 1324 CGTAAGGAGTAGCCCGGAGATTTTAAACCAACCTAGGACCATCGAGAGACCTGTC 1383  
|||||  
Db 490 ArgAlaGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgProVal 509  
|||||  
QY 1384 GCCTACCTGTCAAGAAGCTCGATCTGTAGCCAGTGTGGCCCATATGCTCAAGGCT 1443  
|||||  
Db 510 AlaTyLeuSerLysLysLeuAspProValAlaSerGlyTrpProThrCysLeuLysAla 529  
|||||  
QY 1444 ATCGCAGCTGTGGCCATCTGCTCAAGGACGCTGACAAATGACTTTGGGACAGAAATATA 1503  
|||||  
Db 530 ValAlaAlaValAlaLeuLeuLysAspAlaAspLysLeuThrLeuGlyGlnAsnVal 549  
|||||  
QY 1504 ACTGTATAGCCCCCATGATTTGGAGAACATCGTTGGGAGCCCGACCGATGATG 1563  
|||||  
Db 550 ThrValIleAlaSerHisSerLeuGluSerIleValArgGlnProProAspArgTrpMet 569  
|||||  
QY 1564 ACCAACCGCGCATGACCCACTATCAAGCTGCTCTCACAGAGAGGCTCACCTGCT 1623  
|||||  
Db 570 ThrAsnAlaArgMetThrHisTyGlnSerLeuLeuLeuAsnGluArgValSerPheAla 589  
|||||  
QY 1624 CCACGACCGCTCTCAACCCCTGCCACTTCTTGCTGGAAGAGACTGTAGAACCACTGACT 1683  
|||||  
Db 590 ProProAlaValLeuAsnProAlaThrLeuLeuProValGluSerGluAlaThrProVal 609  
|||||  
QY 1684 CATGATTCCTCAACTATTGATTGAGAGACTGGGTCCCAAGGACCTTACAGACATA 1743  
|||||  
Db 610 HisArgCysSerGluIleLeuAlaGluThrGlyThrArgArgAspLeuGluAspGln 629  
|||||  
QY 1744 CCGCTGACTGAGAAGTGCTAACCTGTTCACTCACGGAAGCAGCTATGTGGTGAAGGT 1803  
|||||  
Db 630 ProLeuProGly---ValProThrTrpTyThrAspGlySerSerPheIleThrGluGly 648  
|||||  
QY 1804 AAGAGATGGTGGGGCGGTGGTGGACGGGACCCGACGATCTGGCGGCAGAGCCTG 1863  
|||||  
Db 649 LysArgArgAlaGlyAlaProIleValAspGlyLysArgThrValTrpAlaSerSerLeu 668  
|||||  
QY 1864 CCGAAGAACTTCAGACACAAAGCTGACTATGCGCTCAGCGACCTTTCGGGCTG 1923  
|||||  
Db 669 ProGluGlyThrSerAlaGlnLysAlaGluLeuValAlaLeuThrGlnAlaLeuArgLeu 688  
|||||  
QY 1924 GCCGAAGGGAATCCATAACATTTATACGACAGCAGGTATGCTTTGCCACTGCACAC 1983  
|||||  
Db 689 AlaGluGlyLysAsnIleAsnIleTyThrAspSerArgTyAlaPheAlaThrAlaHis 708  
|||||  
QY 1984 GTACATGGGGCCATCTATAAACAAGGGGTGCTTACCTCAGCAGGAGGGAATAAAG 2043  
|||||  
Db 709 IleHisGlyAlaIleTyLysGlnArgGlyLeuLeuThrSerAlaGlyLysAspIleLys 728  
|||||  
QY 2044 AACAAAGAGGAAATTTCAAGCCTATTAGAACCGCTACATTTACCAAAAGCGCTAGCTATT 2103  
|||||  
Db 729 AsnLysGluGluIleLeuAlaLeuGluAlaIleHisLeuProArgArgValAlaIle 748  
|||||  
QY 2104 ATACACTGCTCGGACATCACAAAGCTCAATATCTCATATCCAGAGGAAACCATGCT 2163  
|||||  
Db 749 IleHisCysProGlyHisGlnArgGlySerAsnProValAlaThrGlyAsnArgAla 768  
|||||  
QY 2164 CACCGGTTGCCAAGCAGCGCCAGCGGTGTAACTTCTGCTCTATTAATAGAATGCC 2223  
|||||  
Db 769 AspGluAlaAlaLysGlnAlaLeuSerThrArgValLeuAlaGlyThrThrLysPro 788  
|||||  
QY 2224 AAAGCCCA---GAACCCAGACGACGTACACCTTAGAAGACTGGCAGAGATAAAG 2280  
:::|||||  
Db 789 GlnGluProIleGluProAlaGlnGluLysThrArgPro----- 801











```

Db 1027 ThrLysLeuThrLeuAlaThrGlySerArgAspTrpValLeuLeuLeuProLeuAlaLeu 1046
Qy 2956 TTTAGGGTTAGGAACACCCCTGGACAGTTGGGCTGACCCCTATGAATTAATCTACTACGGG 3015
Db 1047 TyrArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGluLeuLeuTyrGly 1066
Qy 3016 GGACCCCCCATTTGGTAGAAATGCTTCTGTACATAGTGTGACGTGCTGCTTCCCG 3075
Db 1067 AlaProProProLeuValAsnPheProAspProAspMetAlaLysValThrHisAsnPro 1086
Qy 3076 CTTTGTCTCTAGGCTCAAGCACTTGTAGTGGGTGACACACAGCGGTGAGGCACTC 3135
Db 1087 SerLeuGlnAlaHisLeuGlnAlaLeuTyrLeuValGlnHisGluValTrpArgProLeu 1106
Qy 3136 CGGAGGCTCTACTCAGGAGGAGGACTTGCAG---ATCCACATCTGTTTCAAGTGGGA 3192
Db 1107 AlaAlaAlaTyrGlnGlnLeuAspArgProValValProHisProPheArgValGly 1126
Qy 3193 GATTCAGTCTAGTTAGACGCCCGTGCAGAACTTCGAGACTCGGTGGAAGGCCCT 3252
Db 1127 AspThrValTrpValArgHisGlnThrLysAsnLeuGluProArgTrpLysGlyPro 1146
Qy 3253 TATCTCGTACTTTGACACACCAACCGCTGTGAAGTTCGAAGAACTCCACCTGGATC 3312
Db 1147 TyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleAlaAlaTrpIle 1166
Qy 3313 CATGCATCCCACTTAACCCGCG-----CCACCTCCGATTCCGGGTGG 3357
Db 1167 HisAlaAlaHisValLysAlaAlaAspThrLysIleGluProProSerGluSerThrTrp 1186
Qy 3358 AAAGCCGAAAGACTGAATAATCCCTTAAGTTCGCTCCATCCG 3402
Db 1187 ArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArg 1201

RESULT 4
S70393
pol polyprotein - Friend murine leukemia virus (strain FB29)
N:Contains: nuclease; proteinase; reverse transcriptase
C:Species: Friend murine leukemia virus
A:Variety: strain FB29
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S70393
R:Perryman, S.; Nishio, J.; Chesebro, B.
Nucleic Acids Res. 19, 6950, 1991
A:Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.
A:Reference number: S70393; MUID:92107687; PMID:1762923
A:Accession: S70393
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: genomic RNA
A:Residues: 1-1204 <PER>
A:Cross-references: UNIPROT:P26809; UNIPARC:UPI0000131P05; EMBL:Z11128; NID:961547; PIDN
A:Experimental source: strain FB29
A:Note: the authors translated the stop codon at position 5 as Gln
C:Genetics:
A:Gene: pol
C:Superfamily: pol polyprotein
C:Keywords: polyprotein; reverse transcriptase

Alignment Scores:
Pred. No.: 7,46-247 Length: 1204
Score: 3809.50 Matches: 720
Percent Similarity: 76.6% Conservative: 160
Best Local Similarity: 62.7% Mismatches: 250
Query Match: 61.6% Indels: 19
DB: 2 Gaps: 10

US-10-723-552-3_COPY_2307_5741 (1-3435) x S70393 (1-1204)

Qy 4 GGTGCCACGGGCACACAGTATCCATGCACTACCCGAGACAGTGTGAGTG 63
Db 56 GlyAlaThrGlyGlyLysArgTyrArgTrpThrThrAspArgValHisLeuAlaThr 75

```

```

Qy 64 GGACGGGTAAACCCACTCGTTTCTGTCATACCTGAGTGCCCGACACCCCTCTTAGGTAGA 123
Db 76 GlyLysValThrHisSerPheLeuHisValProAspCysProTyrProLeuLeuGlyArg 95
Qy 124 GACTTATTGACCAAGATGGGAGACAAATTTCTTTTGA---CAAGGGAAACCGAAGCTG 180
Db 96 AspLeuLeuThrLysLeuLysAlaGlnIleHisPheGluGlySerGlyAlaGlnValVal 115
Qy 181 TCTCAAAATAACAACCTATCCTACCTGTGTGACCCCTCAATTTAGATGACGAATATCGACTA 240
Db 116 GlyProMetGlyGlnProLeuGlnValThrLeuAsnIleGluLysGluTyrArgLeu 135
Qy 241 TACTCTCCCTAGTAAACCTGATCAAAATATA---CAATTTCTGTTGGAAACAGTTTCCC 297
Db 136 HisGluThrSerLysGlyProAspValProLeuGlySerThrTrpLeuSerAspPhePro 155
Qy 298 CAAGCCTGGGACAGAAACCGCAGGATGGTTTGGCAAGCAAGTTCCTCCCAAGATTATT 357
Db 156 GlnAlaTrpAlaGluThrGlyGlyMetGlyLeuAlaValArgGlnAlaProLeuIle 175
Qy 358 CAAGTGAAGCCAGTGCACACCACTGCTCAGTCAGACAGTACCCCTTGAGTAAAGAGCT 417
Db 176 ProLeuLysAlaThrSerThrProValSerIleLysGlnTrpProMetSerGlnGluAla 195
Qy 418 CAAGAAAGGAATTCGGCCGCTATGTCCTCAAGATTAATCCAACAGGCGATCCTAGTTCTCTGTC 477
Db 196 ArgLeuGlyIleLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys 215
Qy 478 CAATCTCCCTGGAAATCTCCCTGCTACCGTTAGAAAGCTGGGACTAATGACTATCGA 537
Db 216 GlnSerProTrpAsnThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg 235
Qy 538 CAGTACAGACTTGTAGAGAGTCAATAAACGGGTGCAGGATATACACCCCAACAGTCCCG 597
Db 236 ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro 255
Qy 598 AACCTTTATAACCTCTTGTGTGCTCTCCACCCCAACGAGAGCTGTATACAGTATTGGAC 657
Db 256 AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp 275
Qy 658 TTAAGAGTCCCTTCTCTGCTGAGATTACACCCCACTAGCCCAACACATTTTTCCTTC 717
Db 276 LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnSerLeuPheAlaPhe 295
Qy 718 GAATGAGAGACTCCAGTACGGAAGAACCGGGCAGCTCACCTGGACCCCGACTCCCCAA 777
Db 296 GluTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln 315
Qy 778 GGGTTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTC 837
Db 316 GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe 335
Qy 838 AGATCCAAACACCTCAGGTGACCTCTCCAGTACGTGATGACCTGCTTTCGGCGGGA 897
Db 336 ArgIleGlnHisProAspLeuIleLeuLeuGlnTyrValAspAspLeuLeuAlaAla 355
Qy 898 GCCACCAACACAGACTCTTAGAAGGCACCAAGCCACTACTGCTGGAATTTGCTGACCTA 957
Db 356 ThrSerGlnLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAspLeu 375
Qy 958 GGCTACAGAGCTCTCTGAAGAGGCCCAAGATTGAGAGAGAGAGGTAAATACTCTGGG 1017
Db 376 GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly 395
Qy 1018 TACAGTTTGGGACCGGCGACGATGCTGACCGAGCGACCGAAGAACTGTAGTCCAG 1077
Db 396 TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGluThrValMetGly 415
Qy 1078 ATACCGGCCCAACACAGCCCAACAAATGAGAGAGTTTTGGGGCAGCTGATTTTGC 1137
Db 416 GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys 435
Qy 1138 AGACTGTGGATCCCGGGTTTGGACCTTAGCAGCCCTTACCGCCCTTACCCAAAGAA 1197

```

Db ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr 455  
QY 1198 AAAGGGGAATTCCTCTGGGCTCTGAGCACAGAGGCAATTGATGCTATCAAAAGGCC 1257  
Db GlyThrLeuPheGluTrpGlyProAspGlnGlnLysAlaTyrGlnGluIleLysGlnAla 475  
QY 1258 CTGCTGACGACCTGCTCTGGCCCTCCTCAGCTAACTAAACCTTTACCTTTATGCTG 1317  
Db LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal 495  
QY 1318 GATGAGCTTAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTAGGACCACTGAGAGA 1377  
Db AspGlnLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTyrParg 515  
QY 1378 CCTGTGCTACCTGTCAAAGACTCGATCCTGTAGCCAGTGGTGGCCCATATGCTG 1437  
Db ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTyrProProCysLeu 535  
QY 1438 AAGGCTATCGAGCTGTGGCCATCTGTCAGGACGCTGACAAATTCGACTTTGGGACG 1497  
Db ArgMetValAlaAlaIleAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGln 555  
QY 1498 AATATACTGTAATAGCCCCCATGCTATTGGAGAACATCGTTGGCGAGCCGCCAGACCGA 1557  
Db ProLeuValIleLeuAlaProHisAlaValGluAlaLeuValLysGlnProProAspArg 575  
QY 1558 TGGATGACCAACGCCCGCATGACCCATATCAAGCCTGCTTCCTC---ACAGAGAGGCTC 1614  
Db TrpLeuSerAsnAlaArgMetThrHisTyrGlnAlaLeuLeuLeuLeuAspThrAspArgVal 595  
QY 1615 ACGTTCGCTCCACAGCCGCTCAACCTCGCCTCTCTGCTGCTGAAGACTGATGAA 1674  
Db GlnPheGlyProIleValAlaLeuAsnProAlaThrLeuLeuPro---LeuProGluGlu 614  
QY 1675 CCAGTCACTCATGATTGCGCATCACTATTGATTAGGAGACTGGGGCTCCGCAAGACCTT 1734  
Db GlyLeuGlnHisAspCysLeuAspIleLeuAlaGluAlaHisGlyThrArgProAspLeu 634  
QY 1735 ACAGACATACCGCTGACTGGAGAAGTGCTAACCTGGTTCTATGACGGAAGCAGCTATGTG 1794  
Db ThrAspGlnProLeuProAspAlaAspHisThrTrpTyrThrAspGlySerSerPheLeu 654  
QY 1795 GTGAGAGTAAGAGATCGCTGGCGGGCTGGTGGACGGGACCGCAGCATCTGGGCC 1854  
Db GlnGluGlyGlnArgLysAlaGlyAlaAlaValThrThrGluThrGluValValTrpAla 674  
QY 1855 AGCAGCTGCCGGAAGGAACCTTCAGCACAAAGGCTGAGCTCATGGCCCTCACGCAAGCT 1914  
Db LysAlaLeuProAlaGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAla 694  
QY 1915 TTGGCGTGGCGGAAGGAATCCATAAACATTTATACGACAGAGATATGCTTTGCG 1974  
Db LeuLysMetAlaGluGlyLysLysLeuAsnValTyrThrAspSerArgTyrAlaPheAla 714  
QY 1975 ACTGCACCTACATGCGGCCATCTATAACAAAGGGGTGCTTACCTGACGAGGAGG 2034  
Db ThrAlaHisIleHisGlyGluIleTyrArgArgGlyLeuLeuThrSerGluGlyLys 734  
QY 2035 GAAATAAAGAACAAAGGAATTCCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGG 2094  
Db GluIleLysAsnLysAspGluIleLeuAlaLeuLeuLysAlaLeuPheLeuProLysArg 754  
QY 2095 CTAGCTATTATACCTGCTCGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAC 2154  
Db LeuSerIleIleHisCysProGlyHisGlnLysGlyAsnArgAlaGluAlaArgGlyAsn 774  
QY 2155 CAGATGGCTGACCGGTTGCCAAGCAGCAGC-----CAGGTTGTTAAC 2199  
Db ArgMetAlaAspGlnAlaArgGluValAlaThrArgGluThrProGluThrSerThr 794  
QY 2200 CTCTGCTTATAATAGAAATGCCAAAGCCCAAGCCAGACGACGACGACGACCTAGAA 2259  
|||||

Db LeuLeuIleGluAsnSerAlaProTyrThrHisGlu---HisPheHisTyrThrValThr 813  
QY 2260 GACTGGCAAGAGATAAAAGATAGACAGTTCTCTGAGACTCCGGAGGACGCTGTAT 2319  
Db AspileLysAspLeuThrLysLeuGlyAlaThrTyrAspAspAlaLys---LysCysTrp 832  
QY 2320 ACCTCAGATGGAGGAATCCTCGCCCAACAAAGAGGGTTAGAAATATGTCCACACAGATA 2379  
Db ValTyrGlnGlyLysProValMetProAspGlnPheThrPheGluLeuLeuAspPheLeu 852  
QY 2380 CATCGTTAAACCCACTAGGAACATAAACACCTGACGAGTTGGTCAGAACATCC----- 2433  
Db HisGlnLeuThrHisLeuSerPheSerLysThrLysAlaLeuLeuGluArgAsnTyrCys 872  
QY 2434 CCTTATCATGTTCTGAGGCTACACAGGAGTGGCTGACTCGGTGGTCAACATGTTGTGCC 2493  
Db ProTyrTyrMetLeuAsnArgAspArgThrLeuLysAspIleThrGluThrCysGlnAla 892  
QY 2494 TGCCAGCTGGTTAATGCTAATCTTCAGAAATGCTCCAGGGAAGAGACTAAGGGGAAGC 2553  
Db CysAlaGlnValAlaAlaSerLysSerAlaValLysGlnGlyThrArgValArgGlyHis 912  
QY 2554 CACCAGCGCTCACTGGGAAGTGCATTCACAGGTAAAGCCGGCTAAATACGGAAC 2613  
Db ArgProGlyThrHisTrpGluIleAspPheThrGluValLysProGlyLeuTyrGlyTyr 932  
QY 2614 AAATACCTATTGTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACTAAG 2673  
Db LysTyrLeuLeuValPheIleAspThrPheSerGlyTrpValGluAlaPheProThrLys 952  
QY 2674 AAAGAGACTTCAACCGTGGTGGCTTAAATAAATACTCGAAGAAATTTTCCAAAGATTGGA 2733  
Db LysGluThrAlaLysValValThrLysLysLeuLeuGluIlePheProArgPheGly 972  
QY 2734 ATACTAAGTAAATAGGTCAGACATGTCACACTTTTGTGTCGCCAGGTAAAGTCAGGA 2793  
Db MetProGlnValLeuGlyThrAspAsnGlyProAlaPheValSerLysValSerGlnThr 992  
QY 2794 CTGGCCAGATATGGGATTTGATTGGAATCTGTAATCTGCATACAGACCCCAAGCTCA 2853  
Db ValAlaAspLeuLeuGlyValAspTrpLysLeuHisCysAlaTyrArgProGlnSerSer 1012  
QY 2854 GGACAGCTAGAGAGATGAATAGAACCATTAAGAGACCCCTTACTATAATTCACCGCGAG 2913  
Db GlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThrLeuAla 1032  
QY 2914 ACTCGGCTTAATGATAGATAGCTCTCTGCGCTTTGTGCTTTTGTAGGTTAGGAACACC 2973  
Db ThrGlySerArgAspTrpValLeuLeuLeuProLeuAlaLeuTyrArgAlaArgAsnThr 1052  
QY 2974 CCTGCACAGTTTGGCTGACCCCTATGNAATTACTCTACGGGGAGCCGCCCATTTGTA 3033  
Db ProGlyProHisGlyLeuThrProTyrGluIleLeuTyrGlyAlaProProLeuVal 1072  
QY 3034 GAAATGCTCTGTACATAGTGTGCTGCTGCTGCCAGCCTTCTCTAGGCTC 3093  
Db AsnPheProAspProAspMetAlaLysValThrHisAsnProSerLeuGluAlaHisLeu 1092  
QY 3094 AAGCAGCTTGTAGTGGTGACAGCAACAGCGGTGGAGGCAACTCCGGGAGGCCCTACTCAGA 3153  
Db GlnAlaLeuTyrLeuValGlnHisGluValTrpArgProLeuAlaAlaAlaTyrGlnGlu 1112  
QY 3154 GGAGAGACTTGCAG---ATCCACATCGTTTCCAAAGTGGAGATTCAGTCTAGCTTGA 3210  
Db GlnLeuAspArgProValValProHisProPheArgValGlyAspThrValTrpValArg 1132  
QY 3211 CGCCACCGTGCAGAAACCTCGAGACTCGGTGGAGGCGCCCTTATCTCGTACTTTTGACC 3270  
Db ArgHisGlnThrLysAsnLeuGluProArgTrpLysGlyProTyrThrValLeuLeuThr 1152  
QY 3271 ACACAAACGGCTGGAAGTCAAGGAATCTCCACCTGGATCCATGTCATCCACGTTAAA 3330  
Db ThrProThrAlaLeuLysValAspGlyIleAlaAlaTrpIleHisAlaAlaHisValLys 1172  
|||||

Qy 3331 CCGCG-----CCACCTCCGATTGCGGTGGAAGCCGAAAGACTGAA 3375  
 Db 1173 AlaAlaaspThrArglleGluProSerGluSerThrTrpArgValGlnArgSerGln 1192  
 Qy 3376 AATCCCTTAGCTCCGCTCCATCGC 3402  
 Db 1193 AsnProLeuIysIleArgLeuThrArg 1201

## RESULT 5

GNMVM7  
 HIV-1 retropepsin (EC 3.4.23.16) - baboon endogenous virus (strain M7)  
 N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
 C;Species: baboon endogenous virus  
 A;Note: host Papio sp. (baboon)  
 C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
 C;Accession: J0261  
 R;Kato, S.; Matsuo, K.; Nishimura, N.; Takahashi, N.; Takano, T.  
 Jpn. J. Genet. 62, 127-137, 1987  
 A;Title: The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genom  
 A;Reference number: J0260  
 A;Molecule type: DNA  
 A;Residues: 1-1189 <KAT>  
 A;Cross-references: UNIPROT:P10272; UNIPARC:UPI0000131EC7; GB:X05470  
 C;Comment: This protein is synthesized as a gag-pol polyprotein.  
 C;Genetics:  
 A;Gene: pol  
 C;Superfamily: pol polyprotein  
 C;Keywords: aspartic proteinase; endonuclease; hydrolase; nucleotidyltransferase; revers  
 F;1-120/Product: retropepsin #status predicted <PTN>  
 F;3-102/Product: retropepsin #status predicted <RT>  
 F;121-797/Product: RNA-directed DNA polymerase #status predicted <REV>  
 F;798-1189/Product: endonuclease #status predicted <EDE>  
 F;27/Active site: Asp (shared with dimeric partner) #status predicted

## Alignment Scores:

Pred. No.: 2,17e-246 Length: 1189  
 Score: 3802.50 Matches: 730  
 Percent Similarity: 76.4% Conservative: 148  
 Best Local Similarity: 63.5% Mismatches: 244  
 Query Match: 61.5% Indels: 27  
 DB: 1 Gaps: 12

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x GNMVM7 (1-1189)

Qy 4 GGTGCCACGGCAACACAGTATCCATGACTACCGAGAACAGTTGACTGGGAGTG 63  
 Db 51 GlyAlaThrGlyArgLysMetHisLysTrpThrAsnArgArgThrValAsnLeuGlyGln 70  
 Qy 64 GGACGGGTAAACCCACTCGTTCTGGTCATACCTGAGTGGCCAGCACCCCTCTTAGGTAGA 123  
 Db 71 GlyMetValThrHisSerPheLeuValProGluCysProTyrProLeuLeuGlyArg 90  
 Qy 124 GACTTATTGACCAAGATGGAGACAAATTCCTTTT---GAACAAGGAAACAGAGTG 180  
 Db 91 AsnLeuLeuThrLysLeuGlyAlaGlnIleHisPheSerGluAlaGlnValLeu 110  
 Qy 181 TCTGCAATTAACAACCTATCATCTGTGTGACCTCCAAATAGATGACGAATATCGACTA 240  
 Db 111 AspArgaspGlyGlnProIleGlnIleLeuThrValSerLeuGlnAspGluHisArgLeu 130  
 Qy 241 TACTCT---CCCTAGTAAAG-----CTGATCAAAATATACAATTCCTGGTTGGAAACAG 291  
 Db 131 PheAspIleProValThrThrSerLeuProAsp-----ValTrpLeuGlnAsp 146  
 Qy 292 TTTCCCAACCTGGGGAGAAACCGCAGGATGGTTTGGCAAGATTCCTCCCAACAA 351  
 Db 147 PheProGlnAlaTrpAlaGluThrGlyLeuGlyArgAlaLysCysGlnAlaProIle 166  
 Qy 352 GTTATTCACTGAAGCCAGTGGCCACACAGTGTCACTGACAGACAGTACCCCTTGAGTAA 411  
 Db 167 IleIleAsnLeuLysProThrAlaValProValSerIleLysGlnTyrProMetSerLeu 186

Qy 412 GAAGCTCAAGAAGGAATTCGGCCGCATGTCCAAAGATTAAATCCCAACAGGCGCATCTAGTT 471  
 Db 187 GluAlaHisMetGlyIleArgGlnHisIleIleLysPheLeuGluLeuGlyValLeuArg 206  
 Qy 472 CTTGTCCTCAATCTCCCTGGAAATATCTCCCTGCTACCGGTTAGAAAGCTGGGACTAATGAC 531  
 Db 207 ProCysArgSerProTrpAsnThrProLeuLeuProValLysProGlyThrGlnAsp 226  
 Qy 532 TATCGACCACTACAGGACTTGAGAGAGTCAATAAACGGGTGCAGGATATACACCCACA 591  
 Db 227 TyrArgProValGlnAsnLeuArgGluIleAsnLysArgThrValAspIleHisProThr 246  
 Qy 592 GTCCCGAACCTTATACCTCTTGTGTGCTCTCCACCCCAACCGAGCTGTGTATACAGTA 651  
 Db 247 ValProAsnProTyrAsnLeuLeuSerThrLeuLysProAspTyrSerTrpTyrThrVal 266  
 Qy 652 TTGACTTAAAGATGCTCTTCTGCTGCTGAGATTACACCCCTACGACCAACCACTTTT 711  
 Db 267 LeuAspLeuLysAspAlaPhePheCysLeuProLeuAlaProGlnSerGlnLeuPhe 286  
 Qy 712 GCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGCGAGCTCACCTCGACCCGACTG 771  
 Db 287 AlaPheGluTrpLysAspProGluArgGlyIleSerGlyGlnLeuThrTrpThrArgLeu 306  
 Qy 772 CCCCAAGGGTTCAAGAACCTCCCGACCATCTTTGACGAAGCCCTTACACAGAGACTCGCC 831  
 Db 307 ProGlnGlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuThr 326  
 Qy 832 AACTTCAGATCCAAACACCTCAGGTAGACCTCTCTCCAGTACGTGGATACCTGCTTCGTG 891  
 Db 327 AspPheArgThrGlnHisProGluValThrLeuLeuGlnTyrValAspAspLeuLeuLeu 346  
 Qy 892 GCGGGAGCCACCAACAGGACTCTTAGAAGGACGACGAGGACACTACTGCTGGGAATCTCT 951  
 Db 347 AlaAlaProThrLysLysAlaCysThrGlnGlyThrArgHisLeuLeuGlnLeuGly 366  
 Qy 952 GACCTAGGCTACAGAGCTCTCTGAAGAGCCGACAGATTTCGAGGAGAGAGGTAAACATAC 1011  
 Db 367 GluLysGlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnThrLysValThrTyr 386  
 Qy 1012 TTGGGTACAGTTTGGGGACGGGACGCGATGCTGACGAGGACGAGGACGAGAAACGTGA 1071  
 Db 387 LeuGlyTyrIleLeuSerGluGlyLysArgTrpLeuThrProGlyArgIleGluThrVal 406  
 Qy 1072 GTCCAGATACCGCCCAACACAGCCCAACAAATGAGAGAGTTTGGGGACAGCTGGA 1131  
 Db 407 AlaArgIleProProProArgAsnProArgGluValArgGluPheLeuGlyThrAlaGly 426  
 Qy 1132 TTTTGACAGACTGTGGATCCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTAAC 1191  
 Db 427 PheCysArgLeuTrpIleProGlyPheAlaGluLeuAlaProLeuTyrAlaLeuThr 446  
 Qy 1192 AAGAAAAAGGGAAATTCCTGGGCTCTCGAGCACAGAGGACATTTGATGTATCAAA 1251  
 Db 447 LysGluSerThrProPheThrTrpGlnThrGluHisGlnLeuAlaPheGluAlaLeuLys 466  
 Qy 1252 AAGGCCCTCTGAGCGACCTGCTGGCCCTCCCTGAGTAACTAAACCTTTACCCCTT 1311  
 Db 467 LysAlaLeuLeuSerAlaProAlaLeuGlyLeuProAspThrSerLysProPheThrLeu 486  
 Qy 1312 TATGTGATGAGCGTAAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTTAGACCATGG 1371  
 Db 487 PheLeuAspGluArgGlnGlyIleAlaLysGlyValLeuThrGlnLysLeuGlyProTrp 506  
 Qy 1372 AGAAGACCTGTGCTTACCTGTCAAAGAGCTCGATCTCTGTAGCCAGTGTGGCCATA 1431  
 Db 507 LysArgProValAlaTyrLeuSerLysLysLeuAspProValAlaGlyTrpProPro 526  
 Qy 1432 TGCCTGAAGGCTATCCAGCTGTGGCCATCTACTGCTCAAGGACGCTCACAAATGACTTTC 1491  
 Db 527 CysLeuArgIleMetAlaAlaThrAlaMetLeuValLysAspSerAlaLysLeuThrLeu 546



```
QY 1492 GGACAGAAATATACTGTAATAGCCGCCCATCGATTGGAGAACATCGTTCCGCGCCGCCCA 1551
Db ||||| :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 547 GlyGlnProLeuThrValIleThrProHisThrLeuGluAlaIleValArgGlnProPro 566
Db ||||| :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1552 GACCGATGGATGACCAACGCCCGCATGACCCCACTATCAAAAGCCTGCTTCTC---ACAGAG 1608
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 567 AspArgTrpIleThrAsnAlaArgLeuThrHisTyGlnAlaLeuLeuLeuAspThrAsp 586
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1609 AGGTTCAGTTCGTCACACAGCGCGCTCAACCCCTGCCACTCTCTCTGCGCTGAAGAGACT 1668
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 587 ArgValGlnPheGlyProProValThrLeuAsnProAlaThrLeuLeuProValProGlu 606
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1669 GATGAACAGTGACTCATGATTGCCATCACTATTGATTGGAGAGACTGGGGTCCGCAAG 1728
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 607 AsnGlnProSerProHisAspCysArgGlnValLeuAlaGluThrHisGlyThrArgGlu 626
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1729 GACCTTACAGACATACCGCTGACTCGGAGAGTGTAACTCGTTCACTGACGGAGAGCAGC 1788
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 627 AspLeuLysAspGlnGluLeuProAspAlaAspHisThrTrpTyThrAspGlySerSer 646
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1789 TATGTGTGGAAAGTAAGAGGATGGCTGGGCGCGGTGGTGGAGCGGCCGCCGACCATC 1848
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 647 TyrLeuAspSerGlyThrArgArgAlaGlyAlaAlaValValAspGlyHisAsnThrIle 666
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1849 TGGGCCAGCGCTGCCGGAAGGAACTTCAGCACAAGAGCGTGAGCTCATCGGCCCTCAGC 1908
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 667 TrpAlaGlnSerLeuProProGlyThrSerAlaGlnLysAlaGluLeuIleAlaLeuThr 686
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1909 CAAGCTTTGGCGGTGGCGGAAGGAAATCCATAAATTTATACGGACAGCAGGTATGCC 1968
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 687 LysAlaLeuGluLeuSerLysGlyLysLysAlaAsnIleTyThrAspSerArgTyAla 706
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1969 TTTCGCACTGCACAGTACATGGGCGCATCTATAAACAAGGGGTGCTTACCTCAGCA 2028
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 707 PheAlaThrAlaHisThrHisGlySerIleTyGluArgArgGlyLeuLeuThrSerGlu 726
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2029 GGGAGGGAATAAGAACAAAGAGGAATTTCTAGCCTATTAGAAGCGGTACATTACCA 2088
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 727 GlyLysGluIleLysAsnLysAlaGluIleAlaLeuLeuLysAlaLeuPheLeuPro 746
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2089 AAAAGGCTAGCTATTATACACTGCTCGCATCAGAAAGCTTAAAGATCTCATATCCAGA 2148
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 747 GlnGluValAlaIleIleHisCysProGlyHisGlnLysGlyGlnAspProValAlaVal 766
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2149 GAAACACAGATGGCTGACCGGTGCCAAGCAGCAGCGCCAGCGGTGTAACTCTGCGCT 2208
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 767 GlyAsnArgGlnAlaAspArgValAlaArgGlnAlaIle-----MetAlaGlu 782
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2209 ATAATAGAAATGCCCAAGCCGCCAGAACCC-----AGACGACAGTAC 2250
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 783 ValLeuThrLeuAlaThrGluProAspAsnThrSerHisIleThrIleGluHisThrTy 802
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2251 ACCCTAGAAAGACTGGCAAGGATAAAAAGATAGACCACTTCTCTGAGACTCCCGAAGGG 2310
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 803 ThrSerGluAspGlnGluAlaArgAlaIleGlyAla---ThrGluAsnLysAspThr 821
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2311 ACCTGCTATACCTCAGATGGGAAGAAATCTGCGCCCAAGAGAGGGTTAGATATGTC 2370
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 822 ArgAsnTrpGluLysGluGlyLysIleValLeuProGlnLysGluAlaLeuAlaMetile 841
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2371 CAACAGATACATCGTCTAACCCACTAGGAACCTAAACACCTGCAGCAGTTCGTCAGRACA 2430
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 842 GlnGlnMetHisAlaTrpThrHisLeuGlyAsnArgLysLeuLysLeuLeuIleGluLys 861
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2431 TCCCTCTATCATGTTCTCAGGCTACCGAGGTGGCTGACTCGGTGGTCAAAACATTGTGTG 2490
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 862 ThrAspPheLeuIleProArgAlaSerThrLeuIleGluGlnValThrSerAlaCysLys 881
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2491 CCCTGCCAGCTGGTTAATGCTTAATCTTCCAGAAATGCTCCAGGAAGAGACTAAGGGGA 2550
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 882 ValCysGlnGlnValAsnAlaGlyAlaThrArgValProAlaAlaGlyLysArgThrArgly 901
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2551 AGCCACCCAGCGGCTCACTGGGAAGTGACTTCACTGAGGTAAAGCCGGCTAAATACGGA 2610
```

```
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 902 AsnArgProGlyValTyTrpGluIleAspPheThrGluValLysProHisTyTrAlaGly 921
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2611 AACAAATACCTATTGGTTTTTTGTAGACACCTTTTTCAGATGGGTAGAGGCTTATCCTACT 2670
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 922 TyrlsTyTrLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaPheProThr 941
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2671 AAGAAAGAGACTTCAACCGTGGTGGCTAAAAAATACTGGAAGAAATTTTCCAGATTTT 2730
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 942 ArgGlnGluThrAlaHisIleValAlaLysLysIleLeuGluIlePheProArgPhe 961
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2731 GGAATACCTAAGGTAATAGGTCAGACAAATGGTCCAGCTTTTGTGGCCAGGTAAAGTCAG 2790
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 962 GlyLeuProLysValIleGlySerAspAsnGlyProAlaPheValSerGlnValSerGln 981
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2791 GGACTGCCCAAGATATTGGGGATTGATGGAAACTGTCATTGTGCATACAGACCCCAAGC 2850
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 982 GlyLeuAlaArgIleLeuGlyIleAsnTrpLysLeuHisCysAlaTyArgProGlnSer 1001
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2851 TCAGGACAGGTAGAGAGATGAATAGAACCAATTAAGAGACCCTTACTTAATTCACCGCG 2910
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1002 SerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThrLeu 1021
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2911 GAGACTGGCGTTAATGATTGGATAGCTCTCTGCGCCCTTTGTGCTTTTAGGGTTAGGAAC 2970
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1022 GluThrGlyLeuLysAspTrpArgArgLeuLeuSerLeuAlaLeuLeuArgAlaArgAsn 1041
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 2971 ACCCTCGACAGTTGGCTGACCCCTATGAATTACTCTACGGGGGACCCCGCCCATTTG 3030
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1042 ThrProAsnArgPheGlyLeuThrProTyGluIleLeuTyGlyGlyProProLeu 1061
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 3031 GTAGAAATGCTTCTGTACATAGTGTGCTGCTGCTTCCACAGCCTTTGTTCTCTAGG 3090
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1062 SerThrLeuLeuAsnSerPheSerProSerAsnSerLysThrAsp---LeuGlnAlaArg 1080
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 3091 CTCAAGGCACCTTGTAGTGGTGAGCAACAGCGGTGGAGGCAACTCCGCGAGGCGCTACTCA 3150
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1081 LeuLysGlyLeuGlnAlaValGlnAlaGlnIleTrpAlaProLeuAlaGluLeuTyArg 1100
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 3151 GGAGGAGAGACTTGCAGATCCCAATCGTTTCAAGTGGGAGATTGATCTAGTCTAGTTAGA 3210
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1101 ProGlyHisSer---GlnThrSerHisProPheGlnValGlyAspSerValTyValArg 1119
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 3211 GCGCACCGTCGAGAAACCTCGAGACTCGGTGGAGGGCCCTTATCTCGTACTTTTGACC 3270
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1120 ArgHisArgSerGlnGlyLeuGluProArgTrpLysGlyProTyIleValLeuLeuThr 1139
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 3271 ACACCAACGGCTGTGAAGTCGAAGGAATCTCCAACCTGGATCCATGCATCCCGCTTAAA 3330
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1140 ThrProThrAlaIleLysValAspGlyIleAlaThrTrpIleHisAlaSerHisAlaLys 1159
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 3331 CCGGCGGCA-----CCTCCGATTCGGGG---TGGAAAGCGGAAAGACTGAA 3375
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1160 AlaAlaProGlyThrProGlyProThrSerSerGlyThrTrpArgLeuArgSerGlu 1179
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 3376 AATCCCTTAAGCTTCGCTCCATCCG 3402
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
QY 1180 AspProLeuLysIleArgLeuSerArg 1188
Db ||||| :||| ||||| ||||| :||| ||||| ||||| :||| ||||| ||||| :||| |||||
```

## RESULT 6

GNMVIM

HIV-1 retropepsin (EC 3.4.23.16) - Moloney murine leukemia virus

N:Contains: nuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA polymera

C:Species: Moloney murine leukemia virus

A:Note: host Mus spp. (mouse)

C:Date: 27-Nov-1985 #sequence\_revision 27-Nov-1985 #text\_change 03-Jun-2002

R:Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.

Nature 293, 543-548, 1981

A:Title: Nucleotide sequence of Moloney murine leukaemia virus.

A:Reference number: A93265; MUID:82035843; PMID:6169994

A:Accession: A03956

A:Molecule type: genomic RNA



A:Residues: 1-1199 <SHI>  
 A:Cross-references: UNIPARC:UPI0000174A30  
 A:Experimental source: clone pMLV-1  
 A:Note: the pol polyprotein contains reverse transcriptase (about 80,000 daltons) and pol yet been defined  
 C:Comment: This protein is synthesized as a gag-pol polyprotein.  
 C:Genetics:  
 A:Gene: pol  
 C:Superfamily: pol polyprotein  
 C:Keywords: aspartic proteinase; hydrolase; nucleotidyltransferase; polyprotein; reverse P;3-102/Product: retropepsin #status predicted <RTP>  
 P;27/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
 Pred. No.: 4.69e-246 Length: 1199  
 Score: 3797.50 Matches: 719  
 Percent Similarity: 76.3% Conservative: 165  
 Best Local Similarity: 62.0% Mismatches: 242  
 Query Match: 61.4% Indels: 33  
 DB: 1 Gaps: 12

US-10-723-552-3\_copy\_2307\_5741 (1-3435) x GNMV1M (1-1199)

```

QY 4 GGTCCGACGGGCAACACAGTATCCATGACTACCCGAGAACAGTGTGGAGTG 63
DB 51 GlyAlaThrGlyGlyLysArgTyrArgTrpThrAspArgLysValHisLeuAlaThr 70
QY 64 GGCGGGTAAACCACTGCTTCTCGTGCATACCTGAGTCCCGACGACCCCTCTTAGGTAGA 123
DB 71 GlyLysValThrHisSerPheLeuHisValProAspCysProTyrProLeuLeuGlyArg 90
QY 124 GACTTATTGACCAAGATGGGAGCACAAATTTCTTTGAA---CAAGGGAAACACGAGAAGTG 180
DB 91 AspLeuLeuThrLysLeuLysAlaGlnIleHisPheGluGlySerGlyAlaGlnValMet 110
QY 181 TCTGCAATTAACAACCTATCAGTGTGTGACCTCCAAATAGATGAGCAATATCGACTA 240
DB 111 GlyProMetGlyGlnProLeuGlnValLeuThrLeuAsnIleGluAspGluHisArgLeu 130
QY 241 TACTCTCCCTAGTAAAGCCGTGATCAAAATATA---CAATTCCTGTTGGAAACAGTTTCCC 297
DB 131 HisGluThrSerLysGluProAspValSerLeuGlySerThrTrpLeuSerAspPhePro 150
QY 298 CAAGCTGGCGAGAAACCGCAGGATGGGTTGGCAAGCAAGATTCCTCCCAAGTTATT 357
DB 151 GlnAlaTrpAlaGluThrGlyGlyMetGlyLeuAlaValArgGlnAlaProLeuIleLe 170
QY 358 CAAGTGAAGCCAGTGCACACCGATGTCAGTCAGACAGTACCCCTTGAGTAAAGAGCT 417
DB 171 ProLeuLysAlaThrSerThrProValSerIleLysGlnTyrProMetSerGlnGluAla 190
QY 418 CAAGAAGGAATTGGCCGCATGTCCTCAAGATTATCAACAGCGCATCTCTAGTTCTGTC 477
DB 191 ArgLeuGlyIleLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys 210
QY 478 CAATCTCCCTGGAAATCTCCCTGCTACCGGTTAGAAAGCCTGGGACTAATGACTATCGA 537
DB 211 GlnSerProTrpAsnThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg 230
QY 538 CCAGTACAGACTTGAGAGAGGTCAATAAACGGGTGCGAGTATACACCAACAGTCCCG 597
DB 231 ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro 250
QY 598 AACCTTATACCTTGTGTGCTCTCCACCCACCGAGCTGGTATACAGTATTGGAC 657
DB 251 AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp 270
QY 658 TTAAGGATGCTTCTCTGCTGAGATTACACCCCACTAGCAACCACTTTTGGCTTC 717
DB 271 LeuLysAspAlaPhePheCysLeuArgLysHisIleProThrSerGlnProLeuPheAlaPhe 290
QY 718 GAATGGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACTGGACCCCGACTGCCCCAA 777

```

```

DB 291 GluTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln 310
QY 778 GGGTTCAAGAACTCCCGACCATCTTTGACGAAGCCCTTACACAGAGACTCGGCCAATTC 837
DB 311 GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe 330
QY 838 AGGATCCAAACACCTCAGTGACCCCTCCTCCAGTACCTGGATGACCTGCTTCTGGCGGA 897
DB 331 ArgIleGlnHisProAspLeuLeuLeuGlnTyrValAspAspLeuLeuAlaAla 350
QY 898 GCCACCAACAGGACTGCTTAGAAGGCACGAGCAGTACTGCTGGAAATTCCTGACCTA 957
DB 351 ThrSerLeuLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAsnLeu 370
QY 958 GGCTACAGAGCCTCTGCTAAGAGGCCCGACAGTTTTCAGGAGAGAGGTAACTACTTGGGG 1017
DB 371 GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly 390
QY 1018 TACAGTTTGGGACGGGACGATGCTGACCGAGGACGAGCAAGAAACTGTAGTCCAG 1077
DB 391 TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGluThrValMetGly 410
QY 1078 ATACCGGCCCAACACACAGCAGCAACAAATCAGAGAGTTTTCGGGACAGCTGGATTTGC 1137
DB 411 GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys 430
QY 1138 AGACTGTGGATCCCGGGTTTGGACCTTAGCAGCCCTACTACCGCTAACCAACAAAGAA 1197
DB 431 ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr 450
QY 1198 AAAGGGGAATTCTCTGGGCTCTCGACCCAGACCAAGGCATTGTATGCTATCAAAAGGCC 1257
DB 451 GlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTyrGlnGluIleLysGlnAla 470
QY 1258 CTGCTGAGCCACCTGCTCTGGCCCTCCCTGAGCTAACTAAACCTTTACCTTTATGTG 1317
DB 471 LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal 490
QY 1318 GATGAGCGTAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTTAGCACCATTGAGAGA 1377
DB 491 AspGlnLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg 510
QY 1378 CTTGCTGCTTACCTGTCAAAGAGCTCGATCTCTGTAGCCAGTGTGTGGCCCATATGCTGT 1437
DB 511 ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTyrProProCysLeu 530
QY 1438 AAGCTATCGCAGCTGTGGCCATCTGCTCAAGACGCTCAACAATTCATCTGGGACAG 1497
DB 531 ArgMetValAlaAlaIleAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGln 550
QY 1498 AATATACTGTAATAGCCCCCATGATTGGAGAACATCGTTCCGACAGCCCGACAGCCGA 1557
DB 551 ProLeuValIleLeuAlaProHisAlaValGluAlaLeuValLysGlnProProAspArg 570
QY 1558 TGGATGACCAACCCCGCATGACCCACTATCAAGCCTGCTTCTC---ACAGAGAGGCTC 1614
DB 571 TrpLeuSerAsnAlaArgMetThrHisTyrGlnAlaLeuLeuLeuAspThrAspArgVal 590
QY 1615 ACCTTCGCTCCACACCGCTCTCAACCCCTGCGACCTCTTCTGCTGAGAGACTGATGAA 1674
DB 591 GlnPheGlyProValValAlaLeuAsnProAlaThrLeuLeuPro---LeuProGluGlu 609
QY 1675 CCAGTCACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGTCCCGACAGGACCTT 1734
DB 610 GlyLeuGlnHisAsnCysLeuAspIleLeuAlaGluAlaHisGlyThrArgProAspLeu 629
QY 1735 ACAGACATACCGCTGACTGGAGAAGTCTAACTGGTTTCACTGCGGAAGCAGCTATGTG 1794
DB 630 ThrAspGlnProLeuProAspAlaAspHisThrTrpTyrThrAspGlySerSerLeuLeu 649
QY 1795 GTGGAAGGTAAAGAGTGTGGCGCGGTGGTGGACGGGACCCGACGATCTGGGCC 1854
DB 650 GlnGluGlyGlnArgLysAlaGlyAlaAlaValThrThrGluThrGluValIleTrpAla 669

```







Db 752 LeuLysAlaThrAlaThrProIleSerIleArgGlnTyrProMetProHisGluAlaTyr 771  
Qy 421 GAAGGAATTCGGCGCATGTCCAAAGATTAAATCAACAGGCGATCCTAGTTCCTGTGCAAA 480  
Db 772 GlnGlyIleLysProHisIleArgMetLeuAspGlnGlyIleLeuLysProCysGln 791  
Qy 481 TCTCCCTGGGAATATCTCCCTGCTACCGGTTAGAAAGCTGGGACTAATGACTATCGACCA 540  
Db 792 SerProTrpAenThrProLeuProValLysLysProGlyThrLysAspTyrArgPro 811  
Qy 541 GTACAGACTTGACAGAGGTCAATAAAGCGGTGCAGGATATACACCAACAGTCCCGAAC 600  
Db 812 ValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValProAsn 831  
Qy 601 CTTTATAACCTCTTGTGTCTCTCCACCCCAACGAGCTGGTATACAGATTTCGACTTA 660  
Db 832 ProTyrAenLeuLeuSerThrLeuProProSerHisProTyrThrValLeuAspLeu 851  
Qy 661 AAGGATCGCTTCTCTGCTGAGATTACACCCCACTAGCCAAACACCTTTTGGCTTCGAA 720  
Db 852 LysAspAlaPhePheCysLeuArgLeuHisSerGluSerGlnLeuLeuPheAlaPheGlu 871  
Qy 721 TGGAGAGATCCAGGTACGGGAAGAACCGGCGAGCTCACCTGGACCCGACTGCCCCAGGG 780  
Db 872 TrpArgAspProGluIleGlyLeuSerGlyGlnLeuThrTrpThrArgLeuProGlnGly 891  
Qy 781 TTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGG 840  
Db 892 PheLysAenSerProThrLeuPheAspGluAlaLeuHisSerAspLeuAlaAspPheArg 911  
Qy 841 ATCCAAACCTCAGGTGACCTCTCCAGTACGTGGATGACCTGCTTCTGGCGGAGCC 900  
Db 912 ValArgTyrProAlaLeuValLeuLeuGlnTyrValAspAspLeuLeuAlaAla 931  
Qy 901 ACCAAACAGGACTCTTAGAGGACCAAGGCACTACTGCTGGNAATGTCTGACCTAGGC 960  
Db 932 ThrArgThrGluCysLeuGluGlyThrLysAlaLeuLeuGluThrLysLysGly 951  
Qy 961 TACAGACCTCTGTAAAGAGGCCAGATTTCAGAGAGAGAGGTAAACATCTTGGGGTAC 1020  
Db 952 TyrArgAlaSerAlaLysLysAlaGlnIleCysLeuGlnGluValThrLysLeuGlyTyr 971  
Qy 1021 AGTTTGGGACGGGACGATGCTGACGGAGGACCGAAGAAACTGTAGTCCAGATA 1080  
Db 972 SerLeuLysAspGlyGlnArgTyrLeuThrLysAlaArgLysGluAlaIleLeuSerIle 991  
Qy 1081 CCGGCCCAACACAGCCAAACAAATCAGAGAGTTTGGGACAGCTGATTTGCAGA 1140  
Db 992 ProValProLysAenProArgGlnValArgGluPheLeuGlyThrAlaGlyTyrCysArg 1011  
Qy 1141 CTGTGGATCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTTAACCAAGAAAAA 1200  
Db 1012 LeuTrpIleProGlyPheAlaGluLeuAlaAlaProLeuTyrProLeuThrArgProGly 1031  
Qy 1201 GGGGAATTCCTGGGCTCTGAGCACCAAGAGGCATTTGATGCTATCAAAAAGGCCCTG 1260  
Db 1032 ThrLeuPheGlnTrpGlyThrGluGlnGlnLeuAlaPheGluAsnIleArgLysAlaLeu 1051  
Qy 1261 CTGAGCGCACTGCTCTGGCCCTCCCTGACGTAACTAAACCTTTACCTTTATGTGGAT 1320  
Db 1052 LeuSerSerProAlaLeuGlyLeuProAspIleThrLysProPheGluLeuPheIleAsp 1071  
Qy 1321 GAGCGTAAGGAGTAGCCGGGAGTTTAAACCAACCCCTAGGACCATGGAGAGACCT 1380  
Db 1072 GluAenSerGlyPheAlaLysGlyValLeuValGlnLysLeuGlyProTrpLysArgPro 1091  
Qy 1381 GTCGCCCTACCTGCAAGAAGCTCGATCTGTAGCCAGTGGTGGGCCCATATGCCTGAAG 1440  
Db 1092 ValAlaTyrLeuSerLysLysLeuAspThrValAlaSerGlyTrpProProCysLeuArg 1111  
Qy 1441 GCTATCGCAGCTGTGGCATACTGCTCAAGGACGTGACAAATGTACTTTGGGACAGAA 1500  
Db 1112 MetValAlaAlaIleAlaIleLeuValLysAspAlaGlyLysLeuThrLeuGlyGlnPro 1131

Qy 1501 ATAACCTGTAATAGCCCCCATGTCATTGGAGAACATCGTTCCGACGCCCCCAGACCGATGG 1560  
Db 1132 LeuThrIleLeuThrSerHisProValGluAlaLeuValArgGlnProProAsnLysTrp 1151  
Qy 1561 ATGACCAACGCCCGCATGACCCACTATCAAAAGCCCTGCTTCTC---ACAGAGAGGGTACG 1617  
Db 1152 LeuSerAenAlaArgMetThrHisTyrGlnAlaMetLeuLeuAspAlaGluArgValHis 1171  
Qy 1618 TTGCTCTCCACGACCGCTCTCAACCCCTGACCTCTCTGCTGAAGAGACTGATGAACCA 1677  
Db 1172 PheGlyProThrValSerLeuAenProAlaThrLeuLeuProLeuProSerGlyLysPro 1191  
Qy 1678 -----GTGACTCATGATTGCCACTATTATTCAGGAGACTGGGTCGCCGAAGGAC 1731  
Db 1192 ProArgLeuSerProAsp-----LeuAlaGluThrMetAlaGlnThrAsp 1206  
Qy 1732 CTTACAGACATACCGCTGACTGGAGAGTCTTAACCTGGTTCACTGACGGAAGCAGCTAT 1791  
Db 1207 LeuThrAspGlnProLeuProAspAlaAspLeuThrTrpTyrThrAspGlySerSerPhe 1226  
Qy 1792 GTGGTGAAGGTAAAGGATGGCTGGCGCGGTGGTGGACGGGACCCGACCATCTGG 1851  
Db 1227 IleArgAenGlyGluArgLysAlaGlyAlaAlaValThrThrGluSerGluValIleTrp 1246  
Qy 1852 GCCAGCAGCTCGCGGAAGGAACTTCAGCACAAAGGCTGAGCTCATGGCCCTCACGCAA 1911  
Db 1247 AlaAlaSerLeuProProGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGln 1266  
Qy 1912 GCTTTGGGCTGCGGAGGAAATCCATAAACATTTATACGACACAGCAGGTATGCCTTT 1971  
Db 1267 AlaLeuLysMetAlaLysLysLysLeuThrValTyrThrAspSerArgTyrAlaPhe 1286  
Qy 1972 GCCACTGCACACTACATGGGCGCCATCTATAACAAGGGGTGCTTACTCTAGCAGGG 2031  
Db 1287 AlaThrAlaHisValHisGlyGluIleTyrArgArgArgGlyLeuLeuThrSerGluGly 1306  
Qy 2032 AGCGAAATAAGAACAAAGAGGAAATCTAAGCCTATTAGAAGCCGTACATTTTACCAAAA 2091  
Db 1307 LysGluIleLysAenLysAenGluIleLeuAlaLeuLeuGluAlaLeuPheLeuProLys 1326  
Qy 2092 AGCTAGCTATTATATACACTGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGA 2151  
Db 1327 ArgLeuSerIleIleHisCysProGlyHisGlnLysGlyAspSerProGlnAlaLysGly 1346  
Qy 2152 AACCATGCTCAGCGGTTGCCAAGCAGCGAGCC-----CAGGGTGTAAAC 2199  
Db 1347 AsnArgLeuAlaAspAspThrAlaLysLysAlaAlaThrGluThrGlnSerSerLeuThr 1366  
Qy 2200 CTTCTGCT-----ATAATAGAAATGCCCAAGCCCAAGCCAGACCCAGACGATACACC 2253  
Db 1367 IleLeuProThrGluLeuIleGluGlyProLys-----ArgProProTrpGluTyrAsp 1384  
Qy 2254 CTAGAAGCTGGCAAGAGATAAAAAATAGACAGTCTCTGAGACTCCCGAAGGGACC 2313  
Db 1385 AspSerAspLeuAspLeuValGlnLysLeuGluAlaHisTyrGluProLysArgGlyThr 1404  
Qy 2314 TGTATATACCTCAGATGGGAAGAAATCCTGCCCCCAAGAAAGGGTTAGAAATATGTCAA 2373  
Db 1405 ---TrpGluTyrArgGlyLysThrIleMetProGluLysTyrAlaLysGluLeuIleSer 1423  
Qy 2374 CAGATACATCGCTAACCCACCTAGGAACCTAAACACTGCAGCAGTTGGTCAGAACATCC 2433  
Db 1424 HisLeuHisLysLeuThrHisLeuSerAlaArgLysMetLysThrLeuLeuGluArgGlu 1443  
Qy 2434 -----CCTTATCATGTTCTGAGGCTACCCAGAGTGGCTGACTCGGTG 2475  
Db 1444 GluThrGlyPheTyrLeuProAenArgAspLeuHisLeuArgGlnValThrGluSer--- 1462  
Qy 2476 GTCAAACATTTGTGCGCTGCCAGCTGTTAATGCTTAATCTCTCAGAAATGCTCTCAGGG 2535  
Db 1463 -----CysArgAlaCysAlaGlnIleAsnAlaGlyLysIleLysPheGlyProAsp 1479



Db 241 ValThrTyrLeuGlyTyrIleLeuSerGluGlyLysArgTrpLeuThrProGlyArgIle 260  
Qy 1063 AAAAAGTGTAGTCCAGATACCGCGCCCAACACACAGCCAAACAAATAGAGAGATTTTGGG 1122  
Db 261 GluThrValAlaHisIleProProGlnAsnProArgGluValArgGluPheLeuGly 280  
Qy 1123 ACAGCTGGATTTTCAGACTGTGATCCCGGGTTTGGACCTTAGCAGCCCACTCTAC 1182  
Db 281 ThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGluLeuAlaAProLeuTyr 300  
Qy 1183 CCGCTAACCAAGAAAGGAATCTCTGGGCTCTGAGCACCAAGAGGCAATTCAT 1242  
Db 301 AlaLeuThrLysGluSerAlaProPheThrTrpGlnGluLysHisGlnSerAlaPheGlu 320  
Qy 1243 GCTATCAAAAGCCCTGCTGAGCGCACCTGCTCTGGCCCTCCCTGACGTAACCTAAACCC 1302  
Db 321 AlaLeuLysGluAlaLeuLeuSerAlaProAlaLeuGlyLeuProAspThrSerLysPro 340  
Qy 1303 TTTACCTTTATGTGGATGAGCGTAAGGAGTAGCCGGGAGTTTAAACCAACCTTA 1362  
Db 341 PheThrLeuPheIleAspGluLysGlnGlyIleAlaLysGlyValLeuThrGlnLysLeu 360  
Qy 1363 GGACCATGGAAGACCTGTGCGCTACCTGTCAAGAGAGCTCGATCTGTAGCCAGTGT 1422  
Db 361 GlyProTrpLysArgProValAlaLysLeuSerLysLysLeuAspProValAlaIleArg 380  
Qy 1423 TGGCCCATATGCTGAAGGCTATCGCAGCTGTGGCCATCTGCTCAAGGACGCTGACAAA 1482  
Db 381 TrpProProCysLeuArgIleMetAlaAlaThrAlaMetLeuValLysAspSerAlaLys 400  
Qy 1483 TTGACTTTGGGACAGATTAACGTGTAATAGCCCCCATGTCATTGGAGAGACATGTTCCG 1542  
Db 401 LeuThrLeuGlyGlnProLeuThrValIleThrProHisAlaLeuGluAlaIleValArg 420  
Qy 1543 CAGCCCCAGACGATGATGACCAACGCGCATGACCCACTATCAAGCCCTGCTTCTC 1602  
Db 421 GlnThrProAspArgTrpIleThrAsnAlaArgLeuThrHisTyrGlnAlaLeuLeu 440  
Qy 1603 ---ACAGAGAGGTTCAGTTTCGCTCCACACGCGCTCTCAACCTGCGCACTCTCTCGCT 1659  
Db 441 AspThrAspArgIleGlnPheGlyProProValThrLeuAsnProAlaThrLeuLeuPro 460  
Qy 1660 GAAGAGACTGATGACCAACGATGACTCATGATGTCATCACTATTGATGAGGAGACTGG 1719  
Db 461 AlaProGluAspGlnGlnSerAlaHisAspCysArgGlnValLeuAlaGluThrHisGly 480  
Qy 1720 GTCGCAAGACCTTACAGACATACCGCTGACTGCGAAGTGTCACTGTTCACTGAC 1779  
Db 481 ThrArgGluAspLeuLysAspGlnGluLeuProAspAlaAspHisSerTrpTyrThrAsp 500  
Qy 1780 GGAAGCAGCTTATGTGGAGGTAAAGAGATGGCTGGCGCGGTGGTGGACGGGACC 1839  
Db 501 GlySerSerTyrIleAspSerGlyThrArgArgAlaGlyAlaAlaValValAspGlyHis 520  
Qy 1840 CGCAGATCTGGCGCAGCAGCTGCGGAGAGAACTTCAGACAAAAGGCTGAGCTCATG 1899  
Db 521 HisIleIleTrpAlaGlnSerLeuProProGlyThrSerAlaGlnLysAlaGluLeuIle 540  
Qy 1900 GCCTCAGCAAGCTTTGCGGTGGCGGAGGGAATCCATAACATTTATACGACAGC 1959  
Db 541 AlaLeuThrLysAlaLeuGluLeuSerGluGlyLysLysAlaAsnIleTyrThrAspSer 560  
Qy 1960 AGGTATGCTTTGCGATGACACAGTATCATGGGCGCATCTATAAAGGGGGTGTGCTT 2019  
Db 561 ArgTyrAlaPheAlaThrAlaHisThrHisGlySerIleTyrGluArgArgGlyLeuLeu 580  
Qy 2020 ACTTCAGCAGGGAGGGAATAAAGAACAAAGAGAGAAATCTTAAGCCCTATTAGAACCGTA 2079  
Db 581 ThrSerGluGlyLysGluIleLysAsnLysAlaGluIleIleAlaLeuLeuLysAlaLeu 600  
Qy 2080 CATTTACCAAAAAGGCTAGCTATTATACATGCTGCGACATCAGAAAGCTAAAGATCTC 2139  
Db 601 PheLeuProArgLysValAlaIleIleHisCysProGlyHisGlnLysGlyLysAspPro 620

Qy 2140 ATATCCAGAGAAACACAGATGGCTGACCGGGTTGCCAAGCAG-----GCAGCCACGGGT 2193  
Db 621 IleAlaThrGlyAsnArgGlnAlaAspGlnValAlaArgGlnValAlaValAlaGluThr 640  
Qy 2194 GTTAACCTTCTGCTATATAATAGAAATGCCCAAGCCCCAGAACCCAGACGACGATACAC 2253  
Db 641 LeuThrLeuThrLysLeuGluGluThrAsnLeuThrThrAsnLysTyrAlaTyrThr 660  
Qy 2254 CTAGACACTGGCAAGAGATAAANAAGATAGACCCAGTTC-----TCTGAGACTCCGAA 2307  
Db 661 ProGluAspGlnGluGluAlaLysAlaIleGlyAlaIleLeuAsnGlnAspThrLysAsp 680  
Qy 2308 GGGACCTGCTATACCTCAGATGGGAAGAAATCTCGCCCAAGAAAGGGTTAGAAATAT 2367  
Db 681 -----TrpGluLysGluGlyLysIleValLeuProArgLysGluAlaLeuAlaMet 697  
Qy 2368 GTCACACAGATACATCGTCTAACCCACTAGGAACCTAACACCTGACGAGCTGGTCAGA 2427  
Db 698 IleGlnGlnMetHisAlaTrpThrHisLeuSerAsnGlnLysLeuLysLeuLeuIleGlu 717  
Qy 2428 ACATCCCTTATCATGTTCTGAGCTACACAGGAGTGGCTGACTCGGTGGTCAACATGT 2487  
Db 718 LysThrAspPheLeuIleProLysAlaGlyThrLeuIleGluGlnValThrSerAlaCys 737  
Qy 2488 GTGCCCTGCCAGCTGGTTAATGCTTAATCTTCCAGAATGCTTCCAGGGAAGAGACTAAGG 2547  
Db 738 LysValCysGlnGlnValAsnAlaGlyAlaThrArgValProGluGlyLysArgThrArg 757  
Qy 2548 GGAAGCCACCCAGCGCTCACTCGGAAGTGGACTTCACTGAGGTAAAGCCGGCTAAATAC 2607  
Db 758 GlyAsnArgProGlyValTyrTrpGluIleAspPheThrGluValLysProHisTyrAla 777  
Qy 2608 GGAACAAATACCTATTGTTTGTAGACACCTTTTCAGATGGGTAGAGGCTTATCCT 2667  
Db 778 GlyTyrLysTyrLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaTyrPro 797  
Qy 2668 ACTAAGAAAGAGACTTCAACCGTGGTGTGCTTAAATAAATACTTGAAGAAATTTTCCAAAG 2727  
Db 798 ThrArgGlnGluThrAlaHisMetValAlaLysLysIleLeuGluGluIlePheProArg 817  
Qy 2728 TTTGGAATACCTAAGGTAAATAGGTCCAGACAAATGGTCCAGCTTTTGTGGCCAGGTAAGT 2787  
Db 818 PheGlyLeuProLysValIleGlySerAspAsnGlyProAlaPheValSerGlnValSer 837  
Qy 2788 CAGGAGTGGCAAGATATTGGGATTTGATTGAAACTGCTTGTGCATACAGACCCCA 2847  
Db 838 GlnGlyLeuAlaArgThrLeuGlyIleAsnTrpLysLeuHisCysAlaTyrArgProGln 857  
Qy 2848 AGCTCAGGACAGGTAGAGAGATCAATAGAACCATTAAGAGACCCCTTACTAAATTGACC 2907  
Db 858 SerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuThr 877  
Qy 2908 GCGGAGACTGCGCTTAAATGATGATGATGCTCTCTGCGCTTTGTGCTTTTGGGTAGG 2967  
Db 878 LeuGluThrGlyLeuLysAspTrpArgArgLeuLeuSerLeuAlaLeuAlaArg 897  
Qy 2968 AACACCCCTGACAGTTTGGCTGACCCCTATGAAATTAATCTTACAGGGGACCCCCCA 3027  
Db 898 AsnThrProAsnArgPheGlyLeuThrProTyrGluIleLeuTyrGlyGlyProProPro 917  
Qy 3028 TTGGTGAATAATCTCTGTACATAGTGTGCTGCTGCTGCTTCCAGCCT----- 3078  
Db 918 LeuSerThrLeuLeuAsnSerPheSerPro-----SerAspProLysThrAsp 933  
Qy 3079 TTGTTCTTAGGTCAAGGCATTTAGTGGTGAGACAAACAGCGGTGAGGCAACTCCGG 3138  
Db 934 LeuGlnAlaArgLeuLysGlyLeuGlnAlaValGlnAlaGlnIleTrpThrProLeuAla 953  
Qy 3139 GAGGCTTACTCAGGAGGAGACTTGCAGATCCCATCGTTTCCAGTGGGAGATTCA 3198  
Db 954 GluLeuTyrArgProGlyHisPro---GlnThrSerTyrProPheGlnValGlyAspSer 972









```
Db 201 LysaspGlnProLeuProGly---ValProAlaTrpTyrThrAspGlySerSerPheIle 219
Qy 1795 GTGGAAGTAAAGAGATGCTGGGCGCGGTGTGGACGGGACCCGACAGATCTGGGCC 1854
Db 220 AlaGluGlyLysArgArgAlaGlyAlaAlaIleValaspGlyLysArgThrValTrpAla 239
Qy 1855 AGCAGCTCGCGGAGGAACTTCAGCACAAGGCTGAGCTCATGGCCCTCACGCAAGCT 1914
Db 240 SerSerLeuProGluGlyThrSerAlaGlnLysAlaGluLeuValAlaLeuThrGlnAla 259
Qy 1915 TTGGCGCTGGCGGAGGAAATCCATAAACATTTATACGACAGCATATGCTTTCGCG 1974
Db 260 LeuargLeuAlaGluGlyArgAspIleAsnIleTyrThrAspSerArgTyrAlaPheAla 279
Qy 1975 ACTGCACGCTACATGGGCGCATCTATAACAAGGGGTTGCTTACCTCAGCAGGGAGG 2034
Db 280 ThrAlaHisIleHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyLys 299
Qy 2035 GAAATAAAGAACAAAGAGGAAATCTTAAGCCTATTAGAAGCGGTACATTTTACCAAAAGG 2094
Db 300 AspIleLysAsnLysGluGluLeuAlaLeuLeuGluAlaIleHisLeuProLysArg 319
Qy 2095 CTAGCTATTATACACTGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAC 2154
Db 320 ValAlaIleIleHisCysProGlyHisGlnLysGlyAsnAspProValAlaThrGlyAsn 339
Qy 2155 CAGATGGCTGACGGGTTGCCAAGCAGCAGCCCGAGGTGTAACTTCTCGCTTATATA 2214
Db 340 ArgargAlaaspGluAlaLysGlnAlaAlaLeuSerThrArgValLeuAlaGluThr 359
Qy 2215 GAAATGCCAAAGCCCAAGACCCAGACCCAGACGACAGTACACCTTAGAAGCTGCCAGAGATA 2274
Db 360 ThrLysPro-----GlnGluLeu 365
Qy 2275 AAAAAGATAGACCACTTCTGAGACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAG 2334
Db 366 IleLysProAlaGlnValLysThrArgPro-----Gly 376
Qy 2335 GAAATCTGCCCCCAAGAAGGGTTAGATATGTCCAACAGATACATCGTCTAACCCAC 2394
Db 377 GluLeuThrProaspArg--GlyLysGluPheIleGlnArgLeuHisGlnLeuThrHis 395
Qy 2395 CTAGGAACCTAACACCTGCAGCAGTTGCTCAGAACATCCCTTATCATGTTCTGAGGCTA 2454
Db 396 LeuGlyProGluLysLeuLeuGlnLeuValAsnArgThrSerLeuLeuIleProAsnLeu 415
Qy 2455 CCAGAGTGGCTGACTCGGTGGTCAAAATTTGTGTGTCCTGCCAGCTGGTTAAATGCTAAT 2514
Db 416 GlnSerAlaValArgGluValThrSerGlnCysGlnAlaCysAlaMetThrAsnAlaVal 435
Qy 2515 CTTTCCAGAAATGCTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGGGCGCTCCTGGGAA 2574
Db 436 ThrThrTyrArgGluThrGlyLysArgGlnArgGlyAspArgProGlyValTyrTrpGlu 455
Qy 2575 GTGACCTTCACTGAGTAAAGCCGCTAAATACCGAACAATACTATTGTTTTCGTA 2634
Db 456 ValAspPheThrGluValLysProGlyArgTyrGlyAsnArgTyrLeuLeuValPheIle 475
Qy 2635 GAC 2637
Db 476 Asp 476
RESULT 12
A42743
pol polyprotein - radiation murine leukemia virus (strain Kaplan) (fragment)
N:Contains: endonuclease (EC 3.1.-.-); proteinase (EC 3.4.21.-); RNA-directed DNA polym
C:Species: radiation murine leukemia virus
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C:Accession: A42743
R:Poliquin, L.; Bergeron, D.; Fortier, J.L.; Paquette, Y.; Bergeron, R.; Rassart, E.
J. Virol. 66, 5141-5146, 1992
A:Title: Determinants of thymotropism in Kaplan radiation leukemia virus and nucleotide
A:Reference number: A42743; MUID:92333703; PMID:1629969
```

```
A:Accession: A42743
A:Molecule type: DNA
A:Residues: 1-581 <POL>
A:Cross-references: UNIPROT:P31795; UNIPARC:UPI0000131P09; GB:M93052; NID:g332065; PIDN:
C:Comment: This protein is likely to be expressed as a fused gag-pol polyprotein.
C:Note: The precise boundary between the mature proteins has not been determined.
C:Genetics:
A:Gene: pol
A:Superfamily: pol polyprotein
C:Keywords: endonuclease; hydrolase; nucleotidyltransferase; polyprotein; reverse trans
Alignment Scores:
Pred. No.: 1.64e-104 Length: 581
Score: 1681.00 Matches: 329
Percent Similarity: 69.9% Conservatives: 84
Best Local Similarity: 55.7% Mismatches: 148
Query Match: 27.2% Indels: 30
DB: Gaps: 6
US-10-723-552-3_COPY_2307_5741 (1-3435) x A42743 (1-581)
Qy 1696 CAACATTGATTGAGGAGACTGGGTCGCGAAGACCTTACAGACATACCCTGACTGGA 1755
Db 2 GluIleLeuAlaGluThrHisGlyThrArgProAspLeuThrAspGlnProIleProasp 21
Qy 1756 GAAGTGCTAACCTGGTTCACCTGACGGAAGCAGCTATGTGTGGAGGTAAAGAGATGGCT 1815
Db 22 AlaAspHisThrTrpTyrThrAspGlySerSerPheLeuGlnGluGlyGlnArgLysAla 41
Qy 1816 GGGCGCGGTGTGGACGGGACCCGACGATCTGGCGCAGCAGCTGCCGGAAGGAACT 1875
Db 42 GlyAlaAlaValThrThrGluThrGluValIleTrpAlaArgAlaLeuProAlaGlyThr 61
Qy 1876 TCAGCACAAAAGCTGAGCTCATGGCTCAGCAAGCTTGGCGCTGGCGAAGGGAAA 1935
Db 62 SerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLys 81
Qy 1936 TCCATAAACATTTATAGGACAGCAGCTATGCTTGGCTGCACAGTACATCGGGCC 1995
Db 82 ArgLeuAsnValTyrThrAspSerArgTyrAlaPheAlaThrAlaHisIleHisGlyGlu 101
Qy 1996 ATCTATAACAAAGGGGTGCTTACCTCAGCAGGAGGAAATAAAGAAACAAAGAGGAA 2055
Db 102 IleTyrLysArgArgGlyLeuLeuThrSerGluGlyArgGluIleLysAsnLysSerGlu 121
Qy 2056 ATTCTAAGCTTATTAGAGCGGTACATTACCAAAAAGCTAGCTATTATATACATGCTCT 2115
Db 122 IleLeuAlaLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysLeu 141
Qy 2116 GGACATCAGAAGCTAAGATCTCATATCCAGAGGAACCAAGTGGCTGACCGGTTGCC 2175
Db 142 GlyHisGlnLysGlyAspSerAlaGluAlaArgGlyAsnArgLeuAlaAspGlnAlaAla 161
Qy 2176 AAGCAGGCGAGCCAGGGGTGTTAACTTCTGCTATAATAGAAATGCCCAAGCCCCAGAA 2235
Db 162 ArgGluAlaAla-----IleLysAlaProPro 170
Qy 2236 CCCAGACAGTACACCTTAGAAGAC----- 2262
Db 171 AspThrSerThrLeuLeuIleGluAspSerThrProTyrThrProAlaTyrPheHisTyr 190
Qy 2263 ---TGGCAGAGATATAAAGATAGACAGTCTCTGAGACTCCCGAAGGAGCC----- 2313
Db 191 ThrGluThrAspLeuLysLysLeuArgGluLeuGlyAlaThrTyrAsnGlnSerGlnGly 210
Qy 2314 TGTATATACCTCAGATGGGAAGAAATCTCGCCCAAGAAAGGGTTAGATAATATGTCCA 2373
Db 211 TyrTrpValPheGlnGlyLysProValMetProAspGlnPheValPheGluLeuLeuAsp 230
Qy 2374 CAGATACATCTCTTAACCCACTAGGAACTAAACACCTGCAGCAGTGGTGGTC-----AGA 2427
Db 231 SerLeuHisArgLeuThrHisLeuGlyTyrGlnLysMetLysAlaLeuLeuAspArgGly 250
```





```

Db      1573 HisCysProGlyHisProArgSerSerGlyIleValGluArgThrAsnArgThrLeuLys 1592
Qy      2887 GAGACCTTACTAAATGACCGGAGAGCTGCGGTTAATGATGGATAGCTCTCTGCGCC 2946
Db      1593 SerLysIleIleLysAlaGlnGluGlnLeuGlnSerLysTrpThrGluValLeuPro 1612
Qy      2947 TTTGTGCTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGCTGACCCCTATGAATTA 3006
Db      1613 TyrValLeuLeuGluMetAlaThrProLysLysHisGlyLeuSerProHisGluIle 1632
Qy      3007 CTCTACGGGGGACCCCGCCCA-----TTGTAGAAATGCTTCTGTACATAGTGCT 3057
Db      1633 ValMetGlyArgProMetLysThrThrTyrLeuSerAspMetSerProLeuTrpAlaThr 1652
Qy      3058 GACGTGCTGCTTTCCAGCCTTTGTCTCTAGGCTCAAGGCATTGAGTGGGTGAGACAA 3117
Db      1653 AspThrLeuValThr-----TyrMetAsn 1660
Qy      3118 CGAGGTGGAGCACTC-----CGGAGGCGCTACTCAGGAGGAGGAGACTTG 3165
Db      1661 LysLeuThrArgGlnLeuSerAlaTyrHisGlnGlnValValAspGlnTrpProSerThr 1680
Qy      3166 CAGATCCACATCGTTTCCAAAGTGGGAGATTTCAGTCTAGCTTACAGCGCCACCGTCAGGA 3225
Db      1681 SerLeuProProGlyProGluProGlySerTrpCysMetLeuArgAsnProLysLys--- 1699
Qy      3226 AACCTCGAGACTCGGTGAAGGCGCTTATCTGCTACTTTTGACACACACCAACCGCTGTG 3285
Db      1700 -----SerSerAsnTrpGluGlyProPheLeuIleLeuLeuSerThrProThrAlaVal 1717
Qy      3286 AAGCTCGAAGAACTCCACCTGATCCATCCATCCACCGTTAA 3330
Db      1718 LysValGluGlyArgProThrTrpIleHisLeuAspHisCysLys 1732

RESULT 14
S12588
pol polyprotein - mink cell focus-forming virus (fragment)
N;Alternate names: integrase
C;Species: mink cell focus-forming virus, MCF
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 26-Aug-1999
C;Accession: S12588
R;Kedar, P.; Khan, A.S.
Nucleic Acids Res. 18, 4022, 1990
A;Title: Nucleotide sequence of the integrase (IN) gene of an endogenous murine leukemia
A;Reference number: S12588; MUID:90326558; PMID:2165259
A;Accession: S12588
A;Molecule type: DNA
A;Residues: 1-409 <KED>
A;Cross-references: UNIPARC:UPI0000028987; EMBL:X52622; NID:950856; PIDN:CAA36849.1; PID
A;Note: the sequences of residues 301-330 and 331-360 are interchanged in the authors' b
C;Superfamily: pol polyprotein
C;Keywords: polyprotein

Alignment Scores:
Pred. No.:      8.04e-69      Length:      409
Score:          1147.00      Matches:      221
Percent Similarity: 70.2%      Conservative: 59
Best Local Similarity: 55.4%      Mismatches: 105
Query Match:     18.6%      Indels:      14
DB:              2          Gaps:          5

US-10-723-552-3_COPY_2307_5741 (1-3435) x S12588 (1-409)
Qy      2248 TACACCTAGAGACTGCGCAAGAGATAAAAGATAGACAGTTCTCTGAGACT----- 2301
Db      14 TyrThrGluThrAsp-----IleLysAsnLeuArgGluLeuGlyAlaThrTyrAsp 30
Qy      2302 CCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATCTCGCCCAACAAGAGGGTTA 2361
Db      31 ArgGluLysLysTyrTrpValLeuGlnGlyLysProValMetProAspGlnPheThrPhe 50
Qy      2362 GAATATGTCACACAGATACATGCTTAACCCACTAGGAACCTAACACCTGCAGCAGTTG 2421

```

RESULT 15  
B46312  
pol polyprotein - human endogenous virus S71

```

Db      51 GluLeuLeuAspPheLeuHisGlnLeuThrHisLeuSerTyrGlnLysMetArgAlaLeu 70
Qy      2422 GTC-----AGAACATCCCTTATCATGTTCTGAGGCTACACGAGAGTGGCTGACTCGGTG 2475
Db      71 LeuAspArgLysGluSerProTyrTrpMetLeuAsnLysAspLysIleLeuHisGluVal 90
Qy      2476 GTCAAACATTTGTGTCCTGCCAGCTGGTTAATGCTTAATCTTCCAGAAATGCTCCAGGG 2535
Db      91 AlaGluSerCysGlnAlaCysValGlnValAsnAlaSerLysAlaLysValGlyProGly 110
Qy      2536 AAGAGACTAAGGGAAGCCACCCAGGCGCTCCTCCTGGGAAGTGGAGCTTCACTGAGTAAG 2595
Db      111 ValArgValArgGlyHisArgProGlyThrHisTrpGluIleAspPheThrGluValArg 130
Qy      2596 CCGGCTAAATACGGAACAAATACCTATTGTTTTTGTAGACACCTTTTTCAGGATGGGTA 2655
Db      131 ProGlyLeuTyrGlyHisLysTyrLeuLeuValPheValAspThrPheSerGlyTrpVal 150
Qy      2656 GAGGCTTATCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAATAATACTGGAAGAA 2715
Db      151 GluAlaPheProThrLysHisGluThrAlaLysValValThrLysLysLeuLeuGlu 170
Qy      2716 ATTTTTCAGATTTGGAATACCTAAGTAATAGGTAGGTGAGCAATGCTCCAGCTTTTCTT 2775
Db      171 IlePheProArgPheGlyMetProGlnValLeuGlyThrAspAsnGlyProAlaPheVal 190
Qy      2776 GCCCAGCTAAGTCAAGGACTGGCCCAAGATATTGGGATTTGGAACCTCATTTGTGCA 2835
Db      191 SerGlnValSerGlnSerValAlaLysLeuLeuGlyIleAspTrpLysLeuHisCysAla 210
Qy      2836 TACAGACCCCAAGCTCAGGACAGGTAGAGAGGATGAATGAACCATTAATAAGAGACCTT 2895
Db      211 TyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThrLeu 230
Qy      2896 ACTAAATGACCGCGGAGACTGGCGTTAATGATTTGATGATGCTCTCTGCGCTTTGTGCTT 2955
Db      231 ThrLysLeuThrLeuAlaThrGlyThrArgAspTrpValLeuLeuLeuProLeuAlaLeu 250
Qy      2956 TTTAGGTGTAGGAACACCCCTGGACAGCTTTGGGCTGACCCCTATGAATTAATCTACGGG 3015
Db      251 TyrArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGluIleLeuTyrGly 270
Qy      3016 GGAACCCCTTGTGTTAGAAATTTGCTTCTGTACATAGTGTGCTGCTCTTTTCCAG 3075
Db      271 AlaProProProLeuValAsnPheHisAspProGluMetSerLysPheThrAsnSerPro 290
Qy      3076 CCTTTGCTCTAGGCTCAAGGACCTTGAAGTGGTGGAGACAAACGAGCGTGGAGGCACTC 3135
Db      291 SerLeuGlnAlaHisLeuGlnAlaLeuGlnAlaValGlnArgGluValTrpLysProLeu 310
Qy      3136 CGGAGGCTTACTCAGGAGGAGGAGACTTGCAG--ATCCACATCGTTTCCAAAGTGGGA 3192
Db      311 AlaAlaAlaTyrGlnAspGlnGlnAspGlnProValIleProPheLeuValGly 330
Qy      3193 GATTCACTCTAGCTTAGACCCCGCTGAGGAAACCTCGAGACTCGGTGGAAGGCGCT 3252
Db      331 AspThrValTrpValArgArgHisGlnThrLysAsnLeuGluProArgTrpLysGlyPro 350
Qy      3253 TATCTCGTACTTTTACCACACCAACCGCTGTGAAAGTCGAAGAAATCTCCACCTGGATC 3312
Db      351 TyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleAlaAlaTrpIle 370
Qy      3313 CATGCATCCCACTTAAACCGCGCCACCTCCCGATTTCGGGG----- 3354
Db      371 HisAlaAlaHisValLysAlaAlaThrThrProProAlaGlyThrAlaSerGlyProThr 390
Qy      3355 TGGAAAGCCGAAAGATCGAAATCCCTTAAGCTTCGCTCCATCGCTGGTTCCT 3411
Db      391 TrpLysValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGlyProPro 409

RESULT 15
B46312
pol polyprotein - human endogenous virus S71

```

C:Species: human endogenous virus S71  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C:Accession: B46312  
R:Werner, T.; Brack-Werner, R.; Leib-Moesch, C.; Backhaus, H.; Erfle, V.; Hehlmann, R.  
Virology 174, 225-238, 1990  
A:Title: S71 is a phylogenetically distinct human endogenous retroviral element with structural similarity to HTLV-1  
A:Reference number: B46312; MUID:90101379; PMID:2152993  
A:Accession: B46312  
A:Molecule type: DNA  
A:Residues: 1-602 <WER>  
A:Cross-references: UNIPROT:Q7LYY5; UNIPARC:UPI000017A87B; GB:M32788  
C:Genetics:  
A:Gene: pol  
C:Keywords: polyprotein

Alignment Scores:  
Pred. No.: 1,79e-68 Length: 602  
Score: 1142.50 Matches: 283  
Percent Similarity: 57.1% Conservative: 75  
Best Local Similarity: 45.1% Mismatches: 203  
Query Match: 18.5% Indels: 67  
DB: 2 Gaps: 18

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x B46312 (1-602)

QY	1378	CTGTGCGCTACCTGTCACAGAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCGCTG	1437
DB	2	ProValIleTyrLeuSerLysArgLeuAspProValAlaSerArgTrpTrpSerCysLeu	21
QY	1438	AAGGCTATCGGAGCTGTGGCCATCTGCTCAAGGACGCTGACAAATTCACATTTGGGACAG	1497
DB	22	TrpAlaIleAlaAlaThrAlaSerLeuIleGlnThrAspLysLeuThrLeuSerGln	41
QY	1498	AATTAACGTGAATAGCCCCCATGTCATGAGAACATCGTTCGGCAGCCCCCAGACCGA	1557
DB	42	AsnLeuThrLeuAlaProArgAlaIleGluThrLeuLeuGlnSerAlaSerGlyLys	61
QY	1558	TGGATGACCAAGCCCGCATGACCCATCATTAAGCTGCTTCACACA--GAGAGGCTC	1614
DB	62	TrpMetSerAsnAlaArgIleLeuGlnTyrGlnSerLeuLeuLeuAspTrpProArgLeu	81
QY	1615	ACGTTGCTGCCACCGCGCTCTCAACCTGCGACTCTTCCTGCTGGAAGAGACATGATGA	1674
DB	82	ThrPheSerProThrArgCysLeuAsnProAlaThrLeuLeuProAspProAspPheThr	101
QY	1675	CCAGTGACTCATGATGTCATCACTATTGATGAGGAGACTGGGGTCCGCAAGACCTT	1734
DB	102	ThrProValHisAspCysGlnGluLeuGluThrThrGluThr--ValArgProAspLeu	120
QY	1735	ACAGACATACCGCTGACTGGAGNAGTGTAACTGTTCACTGACGGAGCAGCTATGTG	1794
DB	121	GlnAspValProLeuLysGluValAspAlaThrValPheThrAspSerSerLeuLeu	140
QY	1795	GTGGAGGTGAAGAGATGCTGGGGCGGGTGGTGGAGCGGACCCGACCATCTGGGCC	1854
DB	141	LysGlnGlyValArgLysAlaGlyAlaAlaValThrMetGluThrAspLysLeuGlnThr	160
QY	1855	AGCAGCTGCCGGAAGAACTTCAGCAAAAGGCTGAGCTCATGGCCCTCAGCGAAGCT	1914
DB	161	GlnAlaLeuProAlaGlyThrSerAlaGlnLysAlaGluLeuValAlaLeuIleGlnAla	180
QY	1915	TTGCGGTGGCGGAGGGAATCCATAAATTTATACGGACAGCAGGTATGCC--TTTGC	1973
DB	181	LeuArgArgVal-ArgThrAsnValLeuThrPheThrLeuThrAlaGlyMetLeuPheAl	200
QY	1974	GACTGCACATGCTGAGGCGCATCTATAAAACAAAGGGGTGTCTTACTCTAGCAGGAG	2033
DB	200	aThrValArgValHisGlyAlaIleTyrGlnValArgGlyLeuLeuThrSerAlaGlyLys	220
QY	2034	GGAAATAAGAACAAAGAGAAATTTCTAAGCTATTAGAAGCCGTACATTTCCAAAAAG	2093
DB	220	SalAileLysAsn--GluGluIleLeuAlaLeuLeuGluAlaValCysLeuProGlnGln	239

QY	2094	GCTAGCTATTATACATCTCTCTGACATCAAGAGCTAAAGATCTCATATCCAGAGGAAA	2153
DB	239	nValAlaValIleHisCysLysGlyHisGlnLysGluAspThrAlaValAlaHisGlyAs	259
QY	2154	CCAGATGGCTGACCGGGTTCGCAAGAGCGGACGCCAG-----GGTGTAACTT	2201
DB	259	nGlnArgAlaAspSerAlaAlaTrpGlyProAlaGlnLeuProValAlaProThrLe	279
QY	2202	TCTGCTTATTAAGAAATGCCAAAGCCCCCAGAACCCAGACGACGATACCCCTAGAGA	2261
DB	279	uLeuProAlaValSerPhePro-----GlnProAspLeuSerAs	292
QY	2262	CTGGCAAGAGATAAAAGATAGACCACTTCTCTGAGACTCCGGAAGGAGCTGCTATAC	2321
DB	292	pHisProGluTyr-----SerProGluGluGlyGlnAl	304
QY	2322	CTCAGATGGGAGGAAATCTGCCCCCAAGAAAGGTTAGAAATATGTCACAGATACA	2381
DB	304	aSerAsp---LeuGlnAlaSerLysAsnGlnGluGly-----	315
QY	2382	TCGTCTAACCCACCTAGGACTTAACACCTGCAGCAGTTGGTCAGAACATCCCTTATCA	2441
DB	316	-----GlyValLys---LeuAlaGlnLeuLeuArg---SerArgPheLys	328
QY	2442	TGTTCTGAGGCTACACGAGGTGGTGTCTCGTGGTCAAACTTGTGTGCTGCTGCAGCT	2501
DB	328	sileProAsnLeuGlnAspLeuValAsnGlnAlaAlaLeuTrpCysThrValCysAlaGl	348
QY	2502	GGTTAATGCTAATCCTTCCAGAATGCCT---CCAGGGAAGAGACTAAGGGGAAGCCACC	2558
DB	348	nValAsnThrLysGlnGlyProLysProSerSerGlyAspArgLeuGlnGlyAspSerPr	368
QY	2559	AGGCGCTCAGTGGGAAGTGGACTTCACCTGAGTAAAGCCGCTAAATACGGAACAATA	2618
DB	368	oGlyGluArgTrpGluIle-----ThrGluIleLysProHisTrpAlaGlyTyrLysTy	386
QY	2619	CCTATTGTTTCTGACACCTTTTCAGGATGGGTAGAGCTTATCTACTATAAGAAAGA	2678
DB	386	rLeuLeuValLeuValAspThrPheSerGly---ThrGluAlaPheAlaThrLysAsnGl	405
QY	2679	GACTTCAACCGTGTGCTAAAAAATACTGGAAGAAATTTTCCAGAGTTTGAATACC	2738
DB	405	uThrAlaThrThrValValLysPheSerLeuAsnGluIleIleProGlnHisGlyLeuPr	425
QY	2739	TAAAGTAAAGGTCAGACATGCTCCAGCTTTTGTGCCAGGTAACTAGTCCAGGACTGC	2798
DB	425	oThrAlaMetGlySerAspAsnArgSerAlaPheThrSerSerIleAlaGlnSerValSe	445
QY	2799	CAAGATATTGGGATTTGATTTGAACTGCATTGTGCATACAGACCCCAAGCTCAGGACA	2858
DB	445	rLysAlaLeuAsnIleGlnTrpLysLeuArgCysAlaTyrArgProGlnSerSerGlyTr	465
QY	2859	GGTAGAGGATGAATAGAACCATTAAGAGACCTTACTAAATTTGACCGCGGAGACTGG	2918
DB	465	pValGluHisMetAsnHisThrLysAsnThrValThrLysLeuIleLeuGluThrGl	485
QY	2919	CGTTAATGATGGATAGCTCTCTCCCTTTGTGCTTTTAGGTTAGGACACCCCTGG	2978
DB	485	yLysAsnGln---ValArgLeuLeuProLeuThrLeuLeuLysValArgCysIleProTy	504
QY	2979	ACAGTTTGGGCTGACCCCTATGAATTAATCTAGCGGGGACCCCTCCATTTGGTA----	3033
DB	504	rArgAlaGlyPheSerProPheGluIleThrTyrArgArgArgProProIleLeuProLy	524
QY	3034	-----GAAATTGCTCTGTACATAGTGTGCTGACGTGCTGCTTTCCCGCCCTTT	3080
DB	524	sLeuLysAspThrArgLeuAlaGluIleSerGluAlaAsnLeuLeu-----GlnTyrLe	542
QY	3081	GTTCTTAGGCTCAGGCACCTTGAGTGGTGAGACACACGCGTGAGGCGCACTCCGGGA	3140
DB	542	uGlnSerLeuGlnValArgAspIleIleGlnProLeuValTrp-----	557
QY	3141	GGCCTACTCAGGAGGAGGAGACTTGCAGATCCCA-----CATCGTTT	3182





**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 12:52:14 ; Search time 142.918 Seconds  
(without alignments)  
3391.447 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_2307\_5741  
Perfect score: 6183  
Sequence: 1 ATGGTGCCACAGGCAACA.....CTGTCAATAACCTCTCAGAC 3435

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSSWEB.spool/US10723552/runat.14022006.125141.12876/app.query.fasta\_1  
-DB-Uniprot -QPMF=fastan -SUFFIX=rup -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USR=US10723552 @CGN 1.1 808 @runat.14022006.125141.12876 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONLOG -DEV TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80.\*  
1: uniprot.sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6026	97.5	1145	2	Q9XSN8_PIG
2	5922.5	96.9	1193	2	Q5QO8_9GAMR
3	5986	96.8	1145	2	Q4VF22_9GAMR
4	5914.5	95.7	1193	2	Q90RL9_9GAMR
5	5885	95.2	2378	2	Q9Q1X4_9GAMR
6	5872	95.0	1720	2	Q8J4V6_9GAMR
7	5866.5	94.9	1718	2	Q8J4V8_9GAMR
8	5864	94.8	2376	2	Q9Q1X5_9GAMR
9	5863	94.8	1146	2	Q8UMP5_9GAMR
10	5861	94.8	1195	2	Q90RL6_9GAMR
11	5861	94.8	2376	2	Q9Q1X3_9GAMR
12	5840.5	94.5	1144	2	Q8UM99_9GAMR
13	5831	94.3	1147	2	Q8UM96_9GAMR
14	5466.5	88.4	1142	2	Q8Q6U4_9GAMR
15	5833.5	87.1	1139	2	Q8Q6U7_9GAMR
16	4450	72.0	868	2	Q73505_9GAMR

17	4156.5	67.2	1127	2	Q9TTC1_PHACI
18	4146.5	67.1	1203	2	O89815_9GAMR
19	4138.5	66.9	1127	2	O70652_GALV
20	4135.5	66.9	1165	1	POL_GALV
21	3835	62.0	1734	1	O7SVK7_9GAMR
22	3833	62.0	1196	1	POL_MLVAV
23	3831	62.0	1733	2	Q9E7M1_9GAMR
24	3825	61.9	1734	2	Q9JBE2_9GAMR
25	3818.5	61.8	1204	1	POL_MLVFP
26	3818	61.7	1196	2	Q90RL4_9GAMR
27	3818	61.7	1734	2	P70355_MOUSE
28	3815.5	61.7	1199	2	Q60FS9_9GAMR
29	3814.5	61.7	1736	2	Q7ZJT6_9GAMR
30	3811.5	61.6	1204	2	O41250_RAU
31	3810.5	61.6	1199	2	Q60FS6_9GAMR
32	3810.5	61.6	1738	2	O39735_MLVFR
33	3809.5	61.6	1204	1	POL_MLVFP
34	3808.5	61.6	1204	2	Q9YK99_9GAMR
35	3802.5	61.5	1189	1	POL_BAEYM
36	3797.5	61.4	1199	1	POL_MLVMO
37	3797.5	61.4	1737	2	O92808_MLVMO
38	3797.5	61.4	1738	2	Q8UN00_MLVMO
39	3785.5	61.2	1204	1	POL_MLVFS
40	3784.5	61.2	1736	2	Q5PYI3_9GAMR
41	3778	61.1	1196	1	POL_MLVRD
42	3772	61.0	1204	2	Q7ZKZ7_9GAMR
43	3757	60.8	1736	2	Q83362_9GAMR
44	3719	60.1	1786	2	O89811_FLV
45	3718.5	60.1	1180	2	Q9WHV7_9GAMR

ALIGNMENTS

RESULT 1  
Q9XSN8\_PIG  
ID Q9XSN8\_PIG PRELIMINARY; PRT; 1145 AA.  
AC Q9XSN8;  
DT 01-NOV-1999 (Trembl)rel. 12, Created  
DT 01-NOV-1999 (Trembl)rel. 12, Last sequence update  
DT 01-MAR-2004 (Trembl)rel. 26, Last annotation update  
DE Pol protein.  
OS Sus scrofa (pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; OC Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=miniature swine;  
RX MEDLINE=98216827; PubMed=9557749;  
RA Akiyoshi D.E., Denaro M., Zhu H., Greenstein J.L., Banerjee P., Fishman J.A.  
RT "Identification of a full-length cDNA for an endogenous retrovirus of miniature swine."  
RL J. Virol. 72:4503-4507(1998).  
DR EMBL; AF038600; AAC16764.1; -, mRNA.  
DR HSP; P03355; 116J.  
DR MEROPS; A02\_020; -.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR001995; Peptidase\_A2\_cat.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF000075; RNaseH; 1.  
DR Pfam; PF000665; rve; 1.  
DR Pfam; PF000077; RVP; 1.

DR Pfam: PF00078; RVT\_1; 1.  
DR PROSITE; PS50175; ISP\_PROT\_RETROV; 1.  
DR PROSITE; PS50879; RNASE\_H; 1.  
SQ SEQUENCE 1145 AA; 128010 MW; 90C1A495B9D95B88 CRC64;  
  
Alignment Scores:  
Pred. No.: 0 Length: 1145  
Score: 6026.00 Matches: 1143  
Percent Similarity: 99.8% Conservative: 0  
Best Local Similarity: 99.8% Mismatches: 2  
Query Match: 97.5% Indels: 0  
DB: 2 Gaps: 0  
  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q9XSN8\_PIG (1-1145)  
  
QY 1 ATGGGTGCCAGGGCAACAAACAGTATCCATGGAGTACCCGAAAGAACAGTTGACTGGGA 60  
DB 1 MetGlyAlaThrGlyGlnGlnLntyrProtrpThrThrArgArgThrValAspLeuGly 20  
  
QY 61 GTGGAGGGTAACCCACTCGTTTCTGTGTACATCTGAGTCCCGACACCCCTCTTAGGT 120  
DB 21 ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 40  
  
QY 121 AGAGACTTATTGACCAAGATGGGAGCAAAATTTCTTTTGAACAAGGGAACACAGAAAGTG 180  
DB 41 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyLysProGluVal 60  
  
QY 181 TCTGGAATAACAAACCTATCACTGTGTGTGACCTCCATTCAGATGAGGAATTCGACTA 240  
DB 61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluLtyrArgLeu 80  
  
QY 241 TACTCTCCCTAGTAAGCCGTGATCAAAATATACAATTTCTGGTTGGACAGTTTCCCAA 300  
DB 81 TyrSerProLeuValLysProAspGlnAsnIleGlnPheTrpLeuGluGlnPheProGln 100  
  
QY 301 GCCTGGGAGAAAACCGAGGAGTGGTTGGCAAGCAAGTTCCCCCAACAAGTTATTCAA 360  
DB 101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 120  
  
QY 361 CTGAAGGCCAGTGCACACCAAGTGTCACTAGACAGATACCCCTTGAGTAAAGAGCTCAA 420  
DB 121 LeuLysAlaSerAlaThrProValSerValArgGlnLtyrProLeuSerLysGluAlaGln 140  
  
QY 421 GAAGGATTCGGCGCATGTCCAAAGATTATCAACAGGGCATCTAGTTCTCTGTCCAA 480  
DB 141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 160  
  
QY 481 TCTCCTGGAAATCTCCCTCTGCTACCGGTTAGAAAGCCTGGGACTAATGACTATCGACA 540  
DB 161 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 180  
  
QY 541 GTACAGACTTGAGAGGTCAATAAACGGGTGACGATATACACCCCAACAGTCCCGAAC 600  
DB 181 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200  
  
QY 601 CCTTATACCTCTGTGTGCTCTCCACCCCACTAGCCCAACCACTTTTGGCTTCGAA 660  
DB 201 ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTrpTyrThrValLeuAspLeu 220  
  
QY 661 AAGGATGCTTCTTCTGCTCAGATTACACCCCACTAGCCCAACCACTTTTGGCTTCGAA 720  
DB 221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240  
  
QY 721 TGGAGAGATCCAGGTACGGGAAGAACCGGGCAGTCTACCTGGACCCGACTGCCCAAGGG 780  
DB 241 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260  
  
QY 781 TTCAGACTCCCGACCATCTTTGACCAAGCCCTACACAGAGACTTGGCCCACTTCAGG 840  
DB 261 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280  
  
QY 841 ATCCAACACCTCTCAGGTGACCCCTCTCCAGTACGTGGATGACCTGCTTCTGGCGGAGCC 900

DB 281 IleGlnHisProGlnValThrLeuLeuGlnLtyrValAspAspLeuLeuLeuAlaGlyAla 300  
QY 901 ACCAAACAGGACTCTCTAGAAGCAGCAGCAAGCACTACTCTGGAAATGTCTGACCTAGGC 960  
DB 301 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 320  
  
QY 961 TACAGAGCCTCTGCTAAGAAGGCCAGATTTCAGAGAGAGAGGTAACATATCTTGGGGTAC 1020  
DB 321 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrLtyrLeuGlyTyr 340  
  
QY 1021 AGTTTGGGACCGGGCAGCGATGGCTGACGAGGACACGGAAGAAACATGTAGTCCAGATA 1080  
DB 341 SerLeuArgAspGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 360  
  
QY 1081 CCGSCCCCAACCCAGCCAAACAATCAGAGAGTTTTTGGGACAGCTGGATTTCGAGA 1140  
DB 361 ProAlaProThrThrAlaLysGlnMetArgGluPheLeuGlyThrAlaGlyPheCysArg 380  
  
QY 1141 CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTAACCCAAAGAAAA 1200  
DB 381 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuLtyrProLeuThrLysGluLys 400  
  
QY 1201 GGGGAATTCCTGGGCTCTGAGCACAGAGGCATTTGATGTATCAAAAAGGCCCTG 1260  
DB 401 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420  
  
QY 1261 CTGAGCCACCTGCTCTGGCCCTCCCTGAGCTAACTAAACCTTTTACCTTTATGTGGAT 1320  
DB 421 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuLtyrValAsp 440  
  
QY 1321 GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCCCAACCCCTAGGACCATGAGAGACCT 1380  
DB 441 GluArgLysGlyValAlaAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 460  
  
QY 1381 GTCCCTTACTGTCAAAAGAGCTCGATCTGTAGCCAGTGGTTGGCCCATATGCTGAAG 1440  
DB 461 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLys 480  
  
QY 1441 GCTATCCAGCTGGGCCATCTGCTCAAGAGCCTGCACAAATTCATCTTCGGGACAGAT 1500  
DB 481 AlaIleAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 500  
  
QY 1501 ATAACTGTATAGCCCCCATGCTTGGAGAAACATCGTTCGGCAGCCCCCAGACCGATGG 1560  
DB 501 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 520  
  
QY 1561 ATGACCAACCGCCGATGACCCACTATCAAGCCTGTCTTCACAGAGAGGTCACGTTTC 1620  
DB 521 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 540  
  
QY 1621 GCTCCACGCGCTCTCAACCCCTGCCACTCTTCTGCTGAAGAGACTGTATGAACCACTG 1680  
DB 541 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 560  
  
QY 1681 ACTCATGATGCCATCAACTATTGATTGAGGAGACTTGGGGTCCGCAAGACCTTACAGAC 1740  
DB 561 ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 580  
  
QY 1741 ATACCGTGACTGAGGAAGTGTAACTGTGTTCTACCGGAAGCAGCTATGTGGTGAA 1800  
DB 581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 600  
  
QY 1801 GGTAAAGGATGGCTGGGGCGGTGGTGCAGCGGACCCGACGATCTGGCCGACGAGC 1860  
DB 601 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 620  
  
QY 1861 CTGCGGAAGAACTTCAGCACAAAAGGCTCAGCTCATGTGCCCTCAGCAAGCTTTGCGG 1920  
DB 621 LeuProGlyGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
  
QY 1921 CTGCGCGAAGGAAATCCATAAACATTATACGACAGCAGGTATGCTTTCGCGACTGA 1980  
DB 641 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 660

QY 1981 CACGTACATGGGGCACTCTATAAACAAGGGGTGGCTTACCTCAGCAGGAGGAGAAATA 2040  
 DB 661 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
 QY 2041 AAGAACAAAGAGGAATTTCTAAGCCTATTAGAAGCCGTACATTTTACCAAAAAGCTAGCT 2100  
 DB 681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaValHisLeuProLysArgLeuAla 700  
 QY 2101 ATTATACACTCTCTGGACATCAGAAAGCTAAGATCTCATATCCAGAGAAACAGATG 2160  
 DB 701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgLysAsnGlnMet 720  
 QY 2161 GCTGACCGGTGGCAAGCAGGACCCAGGGTGTAACTTCTGCCTATAATAAGAAATG 2220  
 DB 721 AlaAspArgValAlaLysGlnAlaGlnGlyValAsnLeuLeuProIleIleGluMet 740  
 QY 2221 CCCAAAGCCCAAGAACCCAGACGACGTACACCTCAGAGACTGGCAGAGATAAAAAG 2280  
 DB 741 ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760  
 QY 2281 ATAGACAGATTCTCTGACACTCCGGAAGGACCTGTATATCTCAGATGGGAAGAAATC 2340  
 DB 761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 780  
 QY 2341 CTGCCCCACAAGAGGGTTAGATATGTCCAACAGATACATCGTCTAACCCACCTAGGA 2400  
 DB 781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
 QY 2401 ACTAAACACTGCAGCAGTGGTGTGAGAACATCCCTTATCATGTTCTGAGGCTACACGGA 2460  
 DB 801 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
 QY 2461 GTGCTCACTCGGTGGTCAACATTTGTGTCCTGCCAGCTGGTAACTGCTAATCCTTCC 2520  
 DB 821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
 QY 2521 AGAATGCTCCAGGAAGAGACTAAGGGGAAGCCACCGCGCTCACTGGGAAGTGGAC 2580  
 DB 841 ArgMetProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860  
 QY 2581 TTCACTGAGTAAGCCGGCTTAATACCGGAACAATACCTATTGTTTGTAGACACC 2640  
 DB 861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
 QY 2641 TTTTCAGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGCTAAA 2700  
 DB 881 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValAlaLys 900  
 QY 2701 AAAATACTGGAAGAAATTTTCCAAAGATTTTGGATACCTAAGGTAAATAGGCTCAGACAAT 2760  
 DB 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
 QY 2761 GTGCCAGCTTTGTTGCCAGGTAAGTCAAGGACTGCGCAAGATATTTGGGATGATGG 2820  
 DB 921 GlyProAlaPheValAlaHisValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
 QY 2821 AAATGCTATTGTGCATACAGACCCCAAGCTCAGGACAGGTAGAGGATGAATAGAAC 2880  
 DB 941 LysLeuHisCysAlaTyrArgProGlnInsSerGlyGlnValGluArgMetAsnArgThr 960  
 QY 2881 ATTAAGAGACCCCTTACTAAATTTGACCGCGAGACTGGCGCTTAATGATTTGGATGCTC 2940  
 DB 961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTrpIleAlaLeu 980  
 QY 2941 CTGCCCCCTTTGCTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGCTGACCCCTAT 3000  
 DB 981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
 QY 3001 GAATTACTCTACGGGGGACCCCGCCCAATTCGTAGAAATTCCTCTGTACATAGTCTGAC 3060  
 DB 1001 GluLeuLeuTyrGlyGlyProProProLeuValGluIleAlaSerValHisSerAlaAsp 1020

QY 3061 GTGCTGCTTTTCCAGCCTTTTCTTAGGCTCAAGGCACTTGTAGTGGGTGAGACAACGA 3120  
 DB 1021 ValLeuLeuSerGlnProLeuPheSerArgLysLysAlaLeuGluTrpValArgGlnArg 1040  
 QY 3121 GCGTGGAGGCAACTCCGGAGGCTTACTCAGGAGGAGAGACTTGCAGATCCCAATCGT 3180  
 DB 1041 AlaTrpArgGlnLeuArgGluAlaTyrSerGlyGlyAspLeuGlnIleProHisArg 1060  
 QY 3181 TTCCAAGTGGGAGATTCACTTACGTTAGACGCCACCGTCAGGAAACCTCGAGACTCGG 3240  
 DB 1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
 QY 3241 TGGAGGCGCCTTATCTCTGTAATTTTGCACACCAACCGCTGTGAAGTCGAAGGAATC 3300  
 DB 1081 TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1100  
 QY 3301 TCCACCTGGATCATGTCATCCAGTTTAAACCGCGCCACCTCCCGATTCGGGTGGAAA 3360  
 DB 1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTrpLys 1120  
 QY 3361 GCCGAAAGACTGAAATCCCTTAAGCTTTCGCTCCATCGCTGGTTCCTTACTCTGTC 3420  
 DB 1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1140  
 QY 3421 AATAACCTCTCAGAC 3435  
 DB 1141 AsnAsnLeuSerAsp 1145  
 RESULT 2  
 Q5QG08\_9GAMR PRELIMINARY; PRT; 1193 AA.  
 ID Q5QG08\_9GAMR PRELIMINARY; PRT; 1193 AA.  
 AC Q5QG08;  
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
 DE Pol protein (Fragment).  
 OS Porcine endogenous retrovirus C/A.  
 OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
 OC Gammaretrovirus; 1-Mammalian type C virus group.  
 OX NCBI\_TaxID=286072;  
 OX [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=A14/220;  
 RX PubMed=15564496; DOI=10.1128/JVI.78.24.13880-13890.2004;  
 RA Bartosch B., Stefanidis D., Myers R., Weiss R., Patience C.,  
 RA Takeuchi Y.;  
 RT "Evidence and consequence of porcine endogenous retrovirus  
 RT recombination.";  
 RL J. Virol. 78:13880-13890(2004).  
 CC -1- SIMILARITY: Contains 1 reverse transcriptase domain.  
 DR EMBL; AV570980; AAT77167.1; -; Genomic DNA.  
 DR GO; GO:0004190; F.aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F.DNA binding; IEA.  
 DR GO; GO:0008233; F.peptidase activity; IEA.  
 DR GO; GO:0004523; F.ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F.RNA binding; IEA.  
 DR GO; GO:0003964; F.RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0006310; P.DNA recombination; IEA.  
 DR GO; GO:0006508; P.proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P.RNA-dependent DNA replication; IEA.  
 DR InterPro; IPR009007; Pept Aspartic cat.  
 DR InterPro; IPR001995; Peptidase\_A2\_cat.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR012337; RNaseH\_fold.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTee.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS50994; INTEGRASE; 1.  
 DR PROSITE; PS50879; RNASE\_H; 1.

DR PROSITE: PS50878; RT POL: 1.  
KW Aspartyl protease; P:cease.  
FT NON\_TER 1  
SQ SEQUENCE 1193 AA; 133310 MW; 9F8EC2AD1A66CE10 CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 1193  
Score: 5992.50 Matches: 1138  
Percent Similarity: 99.7% Conservative: 2  
Best Local Similarity: 99.5% Mismatches: 3  
Query Match: 96.9% Indels: 1  
DB: 2 Gaps: 1  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q50Q8\_9GAMR (1-1193)  
QY 1 ATGGGTGCCACAGGGCAACAACAGTATCCATGAGTACCCGGAAGAACAGTTGACTTGGGA 60  
DB 50 MetGlyAlaThrGlyGlnGlnInfySerTrpThrArgArgThrValAspLeuGly 69  
QY 61 GTGGAGGGTAACCCACTCGTTCTGTCATACCTGAGTCCCGAGCACCCCTCTTAGGT 120  
DB 70 ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 89  
QY 121 AGAGACTTATTGACCAAGATGGAGCACAAAATTTCTTTGAACAGGGAAACCCAGAACTG 180  
DB 90 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyLysProGluVal 109  
QY 181 TCTGCAATAACAACCTTATCACTGTGTGTGACCTCCAAATTAGATGAGCAATATCGACTA 240  
DB 110 SerAlaAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyArgLeu 129  
QY 241 TACTCTCCCTAGTAAGCCTGATCAAAATATACAAATCTGTTGGACAGTTTCCCAA 300  
DB 130 TyrSerProLeuValLysProAspGlnAsnIleGlnPheTrpLeuGluGlnPheProGln 149  
QY 301 GCCTGGGAGAAACCGCAGGAGTGGTTGGCAAGCAAGTTTCCCCCAACAAAGTTATTCAA 360  
DB 150 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 169  
QY 361 CTGAAGCCAGTGGCCACACAGTGTCACTCAGACAGTACCCTGTAGTAAAGAGCTCAA 420  
DB 170 LeuLysAlaSerAlaThrProValSerValArgGlnTyProLeuSerLysGluAlaGln 189  
QY 421 GAAGGAATTCGGCCGATGTCCTCAAGATTATCAACAGGCGATCCTAGTTCTCTGTCAA 480  
DB 190 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 209  
QY 481 TCTCCTCGAATACTCCCTGCTACCGGTTAGAAAGCCTGGGACTTAATGACTATCGACA 540  
DB 210 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyArgPro 229  
QY 541 GTACAGGACTTGAGAGGTCAATAACCGGTGACGATATACACCCCAACAGTCCCGAAC 600  
DB 230 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 249  
QY 601 CCTTATACCTCTGTGTGCTCTCCACCCCAAGAGCTGATACAGTATTGCACTTA 660  
DB 250 ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTrpTyrThrValLeuAspLeu 269  
QY 661 AAGGATGCTTCTTCTGCTGAGATTACACCCCACTAGCCCAACCACTTTTGGCTTCGAA 720  
DB 270 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 289  
QY 721 TGGAGAGATCCAGGTACGGGAAGAACCGGCGAGCTCACCTGAGACCCGACTGCCCAAGGG 780  
DB 290 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 309  
QY 781 TTCAGAACTCCCGACCATCTTTGACCAAGCCCTACACAGAGACCTGGCCAACTTCAGG 840  
DB 310 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 329  
QY 841 ATCCAACACCTCTCAGGTGACCTCTCCAGTACGTGGATGACCTGCTTCTGGCGGAGGCC 900

DB 330 IleGlnHisProGlnValThrLeuLeuGlnTyValAspAspLeuLeuLeuAlaGlyAla 349  
QY 901 ACCAAACAGGACTCTCTAGAAAGGACCAAGCACTACTGCTGGAATTTGCTGACCTAGGC 960  
DB 350 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 369  
QY 961 TACAGACCTCTGCTAAGAAAGGCCAGATTTCAGGAGAGAGAGTAAACATACTCTGGGGTAC 1020  
DB 370 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 389  
QY 1021 AGTTTGGCGGACGGCAGCGATGGCTACCGAGAGCACGGAAGAAACCTGTAGTCCAGATA 1080  
DB 390 SerLeuArgAspGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 409  
QY 1081 CCGCCCCCAACACAGCCCAACAAATGAGAGAGTTTGGGACAGCTGGATTTTGCAGA 1140  
DB 410 ProAlaProThrThrAlaLysGlnMetArgGluPheLeuGlyThrAlaGlyPheCysArg 429  
QY 1141 CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTAACCAAGAAAAA 1200  
DB 430 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLys 449  
QY 1201 GGGGAATTCCTCTGGGCTCTGAGCACAGAGGATTTGATGTATCAAAAAGCCCTG 1260  
DB 450 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 469  
QY 1261 CTGAGCGCACTGCTCTGGCCCTCCCTGACGTAACTAAACCTTTACCTTTATGTGGAT 1320  
DB 470 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyValAsp 489  
QY 1321 GAGCGTAGGAGTAGCCCGGGAGTTTAAACCCCAACCCCTAGGACCATGAGAGAGACCT 1380  
DB 490 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 509  
QY 1381 GTCGCTACCTGTCAAAAGAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCTGAAG 1440  
DB 510 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLys 529  
QY 1441 GCTATCCAGCTGTGGCCATACCTGTCAGGACGCTGACAAATTCACCTTTGGGACAGAT 1500  
DB 530 AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 549  
QY 1501 ATAAGTGTAAATAGCCCCCATGATTCGAGAAACATCGTTCGGCAGCCGCCAGACCGATGG 1560  
DB 550 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 569  
QY 1561 ATGACCAACGCCCATGACCCACTATCAAGCTGTCTTTCACAGAGAGGTCACGTTTC 1620  
DB 570 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 589  
QY 1621 GCTCCACGACGCTCTCAACCCCTGCCACTCTTCTGCTGGAAGACCTGATGAACCCAGTG 1680  
DB 590 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 609  
QY 1681 ACTCATGATTCGCATCAACTATTGATTCAGAGAGACTGGGGTCCGCAAGACCTTACAGAC 1740  
DB 610 ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 629  
QY 1741 ATACCGCTGACTGGAGAGTGTAAACCTGTGTTCTACTGACGGAAGCAGCTATGTGGTGAA 1800  
DB 630 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 649  
QY 1801 GGTAAGAGGATGGCTGGGGGGCGGTGTGACGGGACCCGACAGACTGGCCGAGCAGC 1860  
DB 650 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 669  
QY 1861 CTGCGGGAAGAACTTCAGCAAAAAGCTCAGCTCATGCGCCTCAGCAAGCTTTTGGCG 1920  
DB 670 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 689  
QY 1921 CTGCGCGAAGGAAATCCATAAACATTATACGACAGCAGGTATGCTTTGCCAGCTGCA 1980  
DB 690 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 709

QY 1981 CACGTACATGGGGCCATCTATAAACAAGGGGTGGTCTTACCTCAGCAGGAGGAAATA 2040  
Db HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 729  
QY 2041 AAGAACAAAGAGGAATCTTAAGCCTATTAGAACCGGTACATTACCAAAAAGCTAGCT 2100  
Db LysAsnLysGluGluIleLeuSerLeuLeuGluAlaValHisLeuProLysArgLeuAla 749  
QY 2101 ATTATACACTCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATG 2160  
Db IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 769  
QY 2161 GCTGACCGGGTTGCCAGCAGCGACCCAGGGTGTAACTTCTGCCCTTATATAGAAATG 2220  
Db AlaAspArgValAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProIleGluMet 789  
QY 2221 CCCAAAGCCCGACACCCAGCAGCAGTACACCCCTAGNAGACTGGCAAGAGATAAAAAG 2280  
Db ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTyrGlnGluIleLysLys 809  
QY 2281 ATAGACAGATGTTCTCTGAGACTCCGGAAGGAGCTGCTATACCTCAGATGGGAAGAAATC 2340  
Db IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 829  
QY 2341 CTGCCCCCAAAAGAGGGTTAGAATATGTCCAAACAGATACATCGTCTAAACCCACCTAGGA 2400  
Db LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 849  
QY 2401 ACTAAACACCTGCAGCAGTGGTCAGAACATCCCTTATCATGTTCTGAGGCTACACGGA 2460  
Db ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 869  
QY 2461 GTGCTGACTCGGTGGTCAAACTTGTGTCCTGCCAGCTGGTAAATGCTTAATCTCTCC 2520  
Db ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 889  
QY 2521 AGAATGCTCCAGGAAGAGACTAAGGGGAAGCCACCGAGCGCTCACTGGGAAGTGAC 2580  
Db ArgMetProProGlyLysArgLeuArgLysSerHisProGlyAlaHisTyrGluValAsp 909  
QY 2581 TTGACTGAGTAAAGCCGGCTAAATACGGAACAAATACCTATTGGTTTTGTAGACACC 2640  
Db PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 929  
QY 2641 TTTTCAGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGGTGGCTAAA 2700  
Db PheSerGlyTyrValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 949  
QY 2701 AAAATACCTGGAAGAAATTTTCCAGATTTTGGATACCTAAGGTAAATAGGGTCAGACAAT 2760  
Db LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 969  
QY 2761 GGTCACGTTTGTGGCCAGGTAACTCAGGACTGGCCAGATATTTGGGGATGATGG 2820  
Db GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTyr 989  
QY 2821 AAACCTGCATTGTGCATACAGACCCCAAGCTCAGCAGCAGGTAGAGAGATGAATAGAAC 2880  
Db LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 1009  
QY 2881 ATTAAGAGACCCCTTACTAAATTGACCGCGAGACTGGCGGTAAATGATGGATAGCTCTC 2940  
Db IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTyrIleAlaLeu 1029  
QY 2941 CTGCCCCCTTGTGTTTTAGGGTTAGGAACACCCCTGACAGTGTGGCTGACCCCTAT 3000  
Db LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1049  
QY 3001 GAATTACTCTACGGGGACCCCCCATTCGTAGAAATTCCTCTGTACATAGTCTGAC 3060  
Db GluLeuLeuTyrGlyGlyProProProLeuValGluIleAlaSerValHisSerAlaAsp 1069

QY 3061 GTGCTGCTTTCCACGCTTTTCTTAGGCTCAAGGCACCTTAGTGGGTGAGACAACGA 3120  
Db ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTyrValArgGlnArg 1089  
QY 3121 GCGTGGAGGCACTCCGGAGGCTTACTCAGGAGGAGAGACTTGCAGATCCCATCGT 3180  
Db AlaTyrLysGlnLeuArgGluAlaTyrSer---GlyGlyAspLeuGlnValProHisArg 1108  
QY 3181 TTCCAACTGGGAGATTCACTTACGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGG 3240  
Db PheGlnValGlyAspSerValTyrValArgHisArgAlaGlyAsnLeuGluThrArg 1128  
QY 3241 TGGAGGCGCCCTTATCTCTGACTTTTCACCAACCAACCGCTGTGAAGTCGAAGGAATC 3300  
Db TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1148  
QY 3301 TCACCTTGGATCCATGTCATCCACGTTAAACCGCGCCACCTCCCGATTTCGGGTGAAA 3360  
Db ProThrTyrPheIleHisAlaSerHisValLysProAlaProProAspSerGlyTyrLys 1168  
QY 3361 GCCGAAAGACTGAAAATCCCTTAAGCTTCGCTCCATCGCTGGTTCCTTACTCTGTC 3420  
Db AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1188  
QY 3421 AATAACCTCTCA 3432  
Db AsnAsnSerSer 1192  
RESULT 3  
Q4VFZ2\_9GAMR  
ID Q4VFZ2\_9GAMR PRELIMINARY; PRT; 1145 AA.  
AC Q4VFZ2;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Pol protein.  
GN Namespol;  
OS Porcine endogenous retrovirus C/A.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gammaretrovirus; 1-Mammalian type C virus group.  
OX NCBI\_TaxID=286072;  
RN [1]  
RC NUCLEOTIDE SEQUENCE.  
RA Denner J., Karlas A., Votteler Jr.;  
RT "Nucleotide sequence and extended characterization of a high passage  
human cell-adapted recombinant FERV-C/A.";  
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY953542; AAY28928.1; -; Genomic DNA.  
SQ SEQUENCE 1145 AA; 127925 MW; 62014B9629B04051 CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 1145  
Score: 5986.00 Matches: 1134  
Percent Similarity: 99.6% Conservative: 5  
Best Local Similarity: 99.1% Mismatches: 5  
Query Match: 96.8% Indels: 0  
DB: 2 Gaps: 0  
US-10-723-552-3\_copy\_2307\_5741 (1-3435) x Q4VFZ2\_9GAMR (1-1145)  
QY 1 ATGGGTGCCACAGGCGCAACAACAGTAGTATCCATGAGTACCCGGAAGACAGTTGACTTGGGA 60  
Db 1 MetGlyAlaThrGlyGlnGlnTyrProTyrThrArgThrValAspLeuGly 20  
QY 61 GTGGGACGGGTAAACCACTCGTTTCTGGTCATACCTGAGTCCGACACCCCTCTTAGGT 120  
Db 21 ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 40  
QY 121 AGAGACTTATTGACCAAGATGGGAGGACAAATTTCTTTGAAACAGGAACACAGATG 180  
Db 41 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyLysProGluVal 60

181 TCTGCAATAACAAACCTATCAGTGTGTGACCTCCAAATTAGATGACGAATATCGACTA 240  
Db |||||||  
61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLueu 80  
Qy |||||||  
241 TACTCTCCCTAGTAAAGCCTGATCAAAATATACAAATTCTGGTGGAAACAGTTCCCCCAA 300  
Db |||||||  
81 TyrSerProLeuValLysProAspGlnAsnIleGlnPheTrpLeuGluGlnPheProGln 100  
Qy |||||||  
301 GCCTGGGAGAAACCGCAGAGGATGGTTGGCAAGCAAGTTCCCCCAAGATTATTCAA 360  
Db |||||||  
101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 120  
Qy |||||||  
361 CTGAGGCCAGTGCACACAGTGTCACTCAGACAGTACCCCTGCAGTAAAGAGCTCA 420  
Db |||||||  
121 LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerLysGluAlaGln 140  
Qy |||||||  
421 GAAGGAATTCGGCGCATGTCCTCAAGATTAAATCAACAGGCGATCTAGTTCTCTGTCCAA 480  
Db |||||||  
141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 160  
Qy |||||||  
481 TCTCCTCGGAATACCTCCCTGCTACCGGTTAGAAAGCCTGGGACTAATGACTATCGACCA 540  
Db |||||||  
161 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 180  
Qy |||||||  
541 GTACAGGACTGAGAGGTCATAAACGGGTGCAGGATATACACCAACAGTCCCGAAC 600  
Db |||||||  
181 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200  
Qy |||||||  
601 CTTTATAACCTCTTGTGTGCTCTCCACCCCAACGGAGCTGGTATACAGTATTGGACTTA 660  
Db |||||||  
201 ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTyrTrpThrValLeuAspLeu 220  
Qy |||||||  
661 AAGGATGCTTCTTGTGCTGAGATTACACCCCACTAGCCCAACCACTTTTGGCTTCGAA 720  
Db |||||||  
221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240  
Qy |||||||  
721 TGGAGAGATCCAGGTACGGGAAGACCGGCGAGCTCACCTGGACCCGACTGCCCAAGG 780  
Db |||||||  
241 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260  
Qy |||||||  
781 TTCAGAACTCTCCCGACCATCTTTGACGAAGCCCTTACACAGAGACTTGCCTCAACTCAGG 840  
Db |||||||  
261 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280  
Qy |||||||  
841 ATCCAAACCTCTCAGTGACCTCTCCAGTACGTGGATGACCTGCTTCTGGCGGAGCC 900  
Db |||||||  
281 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla 300  
Qy |||||||  
901 ACCAAACAGGACTGCTTAGAGGCACGAGGCACCTACTGCTGGAATCTCTGACCTAGGC 960  
Db |||||||  
301 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 320  
Qy |||||||  
961 TACAGAGCTCTGCTAAGAAGCCCAAGATTGTCAGGAGAGAGGTAAACATACTTGGGGTAC 1020  
Db |||||||  
321 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 340  
Qy |||||||  
1021 AGTTTGGGAGCGGCAGCGATGGCTGACGAGGACCGAGCAAACTGTAGTCCAGATA 1080  
Db |||||||  
341 SerLeuArgAspGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 360  
Qy |||||||  
1081 CCGGCCCCCAACACAGCAACAAATGAGAGATTTTTGGGACAGCTGGATTTTGCAGA 1140  
Db |||||||  
361 ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 380  
Qy |||||||  
1141 CTGTGGATCCCGGGTTTGGCACTTAGCAGCCCCCACTCTACCCGCTAACCAAGAAAA 1200  
Db |||||||  
381 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLys 400  
Qy |||||||  
1201 GGGGAATTTCTCTGGGCTCTGAGCACCGAAGAGGATTTGATGCTATCAAAAAGCCCTG 1260  
Db |||||||  
401 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420  
Qy |||||||  
1261 CTGAGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCCCTTTATCGTGGAT 1320

421 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 440  
Qy |||||||  
1321 GAGGTTAAGGAGTAGCCCGGGGATTTTAAACCAACCCCTAGGACCATGGAGAGACCT 1380  
Db |||||||  
441 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgPro 460  
Qy |||||||  
1381 GTCCCTACCTGTCAAGAAGCTCGATCCTGTAGCTAGTGTGGCCCATATGCTCGAAG 1440  
Db |||||||  
461 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLys 480  
Qy |||||||  
1441 GCTATCCAGCTGTGGCCCATACTGCTCAAGGACCTGACAAATTGACTTTGGGACAGAA 1500  
Db |||||||  
481 AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 500  
Qy |||||||  
1501 ATAACTGTATAGCCCCCATGTCATTGGAGAAATCCTGTTGGGAGCCGCCAGACCGATGG 1560  
Db |||||||  
501 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 520  
Qy |||||||  
1561 ATGACCAACCGCCGCTCTCAACCCCTGCCACTCTTCTGCTGAAAGACTGTAAACCACTG 1620  
Db |||||||  
521 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuThrGluArgValThrPhe 540  
Qy |||||||  
1621 GCTCCACGACGCGCTCTCAACCCCTGCCACTCTTCTGCTGAAAGACTGTAAACCACTG 1680  
Db |||||||  
541 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGlnThrAspGluProVal 560  
Qy |||||||  
1681 ACTCATGATTGCCATCAACTATTGATTGAGAGACTGGGTCCGCAAGGACCTTACAGAC 1740  
Db |||||||  
561 ThrHisaspCysHisGlnLeuLeuIleGluThrGlyValArgLysAspLeuThrAsp 580  
Qy |||||||  
1741 ATACCGCTGACTGAGAAAGTGTAACTGGTTCACTGACGGAAGCAGTATGTGTGGAA 1800  
Db |||||||  
581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 600  
Qy |||||||  
1801 GGTAAAGGATGGTGGCGCGGTGTGTGACGGGACCCCGCAGATCTGCGCCAGCAGC 1860  
Db |||||||  
601 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 620  
Qy |||||||  
1861 CTGCGGGAAGAACTTCAGCACAAAAGCTGAGCTCATGGCCCTCAGCAAGCTTTGGCG 1920  
Db |||||||  
621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
Qy |||||||  
1921 CTGCGCCGAAGGAATCCATAAACATTTATACGACAGCAGGTATGCTTTCGCACTGCA 1980  
Db |||||||  
641 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 660  
Qy |||||||  
1981 CACGTACATGGGCGCATCTATAAAACAAAGGGGTGTGCTTACCTCAGCAGGAGGGAATA 2040  
Db |||||||  
661 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
Qy |||||||  
2041 AAGAACAAGAGGAAATCTTAAGCCTATTAGAAGCTGATCAATTTACCAAAAAGCGTAGCT 2100  
Db |||||||  
681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 700  
Qy |||||||  
2101 ATTATACACTGCTCTGGACATCAGAAAGTCTCATATCCAGAGGAACCCAGATG 2160  
Db |||||||  
701 IleIleHisCysProGlyHisGlnLysAlaLysAspProIleSerArgGlyAsnGlnMet 720  
Qy |||||||  
2161 GCTCAGCGGGTGGCAAGCAGCGCCAGGGGTGTAACTTCTGCTTATATAGAAATG 2220  
Db |||||||  
721 AlaAspArgValAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProMetIleGluThr 740  
Qy |||||||  
2221 CCCAAAGCCCGAAGCCAGACGACGTACACCCCTAGAGACTGGCAGAGATTAAGAAAG 2280  
Db |||||||  
741 ProlysAlaProGluProGlyArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760  
Qy |||||||  
2281 ATAGACCAAGTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGGAAATC 2340  
Db |||||||  
761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 780  
Qy |||||||  
2341 CTGCGCCCAAGAAGGGTTAGAAATATGTCCAACAGATACATCGTCTTAACCCCACTAGGA 2400  
Db |||||||

```
Db      781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800
QY      2401 ACTAAACCTGCAGCAGTGGTTCAGAAATCCCTTCATGTTCTCAGGCTACCGGA 2460
Db      801 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820
QY      2461 GTGGCTGACTCGGTGGTCAAACTGTGTGCGCTGCCAGCTGGTTAATGCTTAACCTTCC 2520
Db      821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840
QY      2521 AGAATGCCTCCAGGGAAGAGACTAAGGGGAAGACACCCAGCGCTCACTGGGAAGTGAC 2580
Db      841 ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860
QY      2581 TTCCTAGGTAAGCCCGCTAAATACGGAACAATAACCTATTGGTTTTGTAGACACC 2640
Db      861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880
QY      2641 TTTTCAGGATGGTAGAGCTTATCCTACTAAGAAGAGACTTCACCGTGGTGTAA 2700
Db      881 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900
QY      2701 AAAATACCTGGGAAGAAATTTTCCAGATTTCGAATACCTAAGGTAATAGGTCAGACAT 2760
Db      901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920
QY      2761 GGTCCAGCTTTTGTGGCCAGCTAAGTCAGGAGCTGGCCAAAGATATTGGGATGATTGG 2820
Db      921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940
QY      2821 AAATGCTGATTGTGCATACAGACCCCAAGCTCAGACAGGTAGAGAGATGAATAGAAC 2880
Db      941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960
QY      2881 ATTTAAAGACCTTACTTAATTCACCGCGAGACTGCGGTAATGATTGGATAGCTCTC 2940
Db      961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTrpIleAlaLeu 980
QY      2941 CTGCCCTTTGTGCTTTTGTAGGGTTAGGAACACCCCTGCAGACAGTTGGGCTGACCCCTAT 3000
Db      981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000
QY      3001 GAATTACTCTACGGGGAGACCCCTTCTAGGCTCAAGGCTTGAATGCTTCTGTACATAGTCTGAC 3060
Db      1001 GluLeuLeuTyrGlyGlyProProProLeuValGluIleAlaSerValHisSerAlaAsp 1020
QY      3061 GTGCTGCTTTCACGACCTTTGTTCTCTAGGCTCAAGGCTTGAATGCTTCTGTACATAGTCTGAC 3120
Db      1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040
QY      3121 CGCTGGAGGCAACTCCGGGGAGCTTACTCAGGAGGAGAGACTTGCAGATCCCACTCGT 3180
Db      1041 AlaTrpArgGlnLeuArgGluAlaTyrSerGlyGlyAspLeuGlnIleProHisArg 1060
QY      3181 TTCCAAGTGGAGATTGAGTCTAGTACGCCACCGCTGAGGAAACCTCGAGACTCGG 3240
Db      1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080
QY      3241 TGGAGGCGCTTATCTCGTACTTTTGCACACCAACCGCTGGAAGCTGGAAGGATC 3300
Db      1081 TrpLysGlyProTyrHisValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1100
QY      3301 TCCACTGGATCCATCCAGTCCAGCTTAAACCGCGCCACTCCCGGATTCGGGGTGAAA 3360
Db      1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTrpLys 1120
QY      3361 GCCGAAAGACTGAAATTCCTTAAAGCTTGCCTCCATCGCTGCTTCTTACTCTGTC 3420
Db      1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1140
QY      3421 AATAACCTCTCA 3432
Db      1141 AsnAsnPheSer 1144
```

```
RESULT 4
Q90RL9_9GAMR
ID Q90RL9_9GAMR PRELIMINARY; PRT; 1193 AA.
AC Q90RL9_
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polymerase (Fragment).
GN Name=p01;
OS Porcine endogenous type C retrovirus.
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC Gammaretrovirus.
OX NCBI_TaxID=70540;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21256017; PubMed=11356953;
RX DOI=10.1128/JVI.75.12.5465-5472.2001;
RA Krach U., Fischer N., Czaderna F., Toenjes R.R.;
RT "Comparison of replication-competent molecular clones of porcine
RT endogenous retrovirus class A and class B derived from pig and human
RT cells.";
RL J. Virol. 75:5465-5472(2001).
CC -; SIMILARITY: Contains 1 reverse transcriptase domain.
DR EMBL; AJ293656; CAC39617.1; -; Genomic_DNA.
DR HSSP; P03355; 116J.
DR MEROPS; A02.020; -.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR009007; Pept_Aspartc_cat.
DR InterPro; IPR001995; Peptidase_A2_cat.
DR InterPro; IPR001237; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT; 1.
DR PROSITE; PS01075; ASP_PROT_RETROV; 1.
DR PROSITE; PS00994; INTEGRASE; 1.
DR PROSITE; PS00879; RNASE_H; 1.
DR PROSITE; PS00878; RT_POL; 1.
DR KW Aspartyl protease; protease.
FT NON_TER 1
SQ SEQUENCE 1193 AA; 133225 MW; 8FC1F419CB600332 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1193
Score: 5914.50 Matches: 1119
Percent Similarity: 98.9% Conservative: 12
Best Local Similarity: 97.8% Mismatches: 12
Query Match: 95.7% Indels: 1
DB: 2 Gaps: 1

US-10-723-552-3_COPY_2307_5741 (1-3435) x Q90RL9_9GAMR (1-1193)
QY 1 ATGGGTGCCAGCGGCGCAACAGTATCCATGACTACCCGAGAACAGTTGACTTGGGA 60
Db 50 MetGlyAlaThrGlyGlnGlnTyrProTyrThrArgArgThrValAspLeuGly 69
QY 61 GTGGGCGGTAAACCCACTCGTTTCTGGTCATACCTGAGTCCAGCACCCCTCTTAGGT 120
Db 70 ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 89
QY 121 AGAGACTTATTGACCAAGATGGGAGCACAATTTCTTTTGAACAGGGAACCAAGATG 180
```



Db 90 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyLysProGluVal 109  
Qy 181 TCTGCAATAACAAACCTATCACTGTGTGTGACCTCCAAATTTAGATGACGAATATCGACTA 240  
Db 110 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyArgLeu 129  
Qy 241 TACTCTCCCTAGTAGAAGCCTGATCAAAATATACAAATTCCTGTTGGACAGTTTCCCCAA 300  
Db 130 TyrSerProLeuValLysProAspGlnAsnIleGlnPheTrpLeuGluGlnPheProGln 149  
Qy 301 GCCTGGGCAGAAACCGCAGGATGGTTTGGCAAGCAAGTTCCCCCAAGATTATTCAA 360  
Db 150 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 169  
Qy 361 CTGAAGCCAGTGCCACACCAAGTGTCACTGACAGACAGTACCCCTTGAGTAAAGAGCTCAA 420  
Db 170 LeuLysAlaSerAlaAlaProValSerValArgGlnTyrProLeuSerLysGluAlaArg 189  
Qy 421 GAAGGAATTCGCGCATGTCCTCAAGATTATCCACAGGGCATCTAGTTCTCTGTCCAA 480  
Db 190 GluGlyIleArgProHisValArgLeuIleGlnGlnGlyIleLeuValProValGln 209  
Qy 481 TCTCCTCGAATACTCCCTCTGCTACCGGTAGAAAGCCTGGGACTTAATGACTATCGACCA 540  
Db 210 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 229  
Qy 541 GTACAGGACTTCAGAGAGTCAATAAACCGGTGACAGATATACACCCCAAGTCCCGAAC 600  
Db 230 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 249  
Qy 601 CCTTATAACCTCTGTGTGCTCTCCACCCCAACCGGACTGTATACAGTATTGGACTTA 660  
Db 250 ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTrpTyrThrValLeuAspLeu 269  
Qy 661 AAGGATGCTTCTTCTGCTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTCGAA 720  
Db 270 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 289  
Qy 721 TGGAGAGATCCAGTACGGGAAGAACCGGGCAGTCTACCTGGACCCGACTCCCCCAAGG 780  
Db 290 TrpArgAspProGlyAlaGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 309  
Qy 781 TTCAGAACTCCCGACCATCTTTGACCAAGCCCTACACAGACCTGGCCCACTTCAGG 840  
Db 310 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 329  
Qy 841 ATCCAAACACCTCAGTGACCCCTCTCCAGTACGTGGATGACCTGCTTCTGGCGGAGCC 900  
Db 330 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla 349  
Qy 901 ACCAAACAGGACTGCTTAGAGGACGAAAGGCACTACTGCTGGAATTCCTGACCTAGGC 960  
Db 350 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 369  
Qy 961 TACAGAGCTCTGCTAAGAGCCAGATTTTCAGGAGAGGTTAACAATCTGGGTAC 1020  
Db 370 TyrArgAlaSerAlaLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 389  
Qy 1021 AGTTTGGGGAGGGCAGCGATGGCTGACGGAGGCACGAAAGAACTGTAGTCCAGATA 1080  
Db 390 SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysArgThrValValGlnIle 409  
Qy 1081 CCGGCCCCAACACACGCAAAATAGAGAGATTTTTGGGACAGCTGGATTTGCGAGA 1140  
Db 410 ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 429  
Qy 1141 CTGTGGATCCCGGGTTTGGACCTTAGACGCCCACTCTACCCGCTTAACCAAGAAAA 1200  
Db 430 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGlnLys 449  
Qy 1201 GGGGAATTCCTCTGGGCTCCTGAGCACGAGGATTTTGATGCTTATCAAAAAGGCCCTG 1260

Db 450 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 469  
Qy 1261 CTGAGCCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCTTTTACCTTTATGTGGAT 1320  
Db 470 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 489  
Qy 1321 GAGCGTAAGGGAGTAGCCCGGGAGTTTAAACCCCAACCTCTAGGACCATGAGAGACCT 1380  
Db 490 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 509  
Qy 1381 GTCCCTACCTGTCAAAAGAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCTGAAG 1440  
Db 510 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTyrProIleCysLeuLys 529  
Qy 1441 GCTATCCAGCTGTGGCCATCTGGTCAAGGACGCTGACAAATTCAGCTTTGGGACAGAT 1500  
Db 530 AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 549  
Qy 1501 ATAACTGTATAGCCCCCATGCAATTCGAGAAACATCTTCGGCAGCCGCCAGACCGATGG 1560  
Db 550 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 569  
Qy 1561 ATGACCAACGCCCATGACCCACTATCAAAGCCTGCTTCTCACAGAGAGGTCACTTC 1620  
Db 570 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 589  
Qy 1621 GCTCACAGCGCTCTCAACCTGCCCCTGCTTCTGCTGAGAGACTGATGACCACTG 1680  
Db 590 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 609  
Qy 1681 ACTCATGATTCGCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGAC 1740  
Db 610 ThrHisaspCysHisGlnLeuLeuIleGluThrGlyValArgLysAspLeuThrAsp 629  
Qy 1741 ATACCGCTGACTGAGGAAGTGTCTAACCTGGTTCTACTACGGAACGAGCTATGTGGTGA 1800  
Db 630 IleProLeuThrGlyGluMetLeuThrTrpPheThrAspGlySerSerTyrMetValGlu 649  
Qy 1801 GGTAAAGAGATGGCTGGGGCGGGTGGTGGACGGGACCCGACGATCTGGGCGGACGAC 1860  
Db 650 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 669  
Qy 1861 CTGCGCGAAGGAACCTTCAGCACAAAGCTGAGCTCATGGCCCTCAGCAAGCTTTGGCG 1920  
Db 670 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 689  
Qy 1921 CTGCGCGAAGGAATAATCCATAAACATTATATACGACGACGAGTATGCTTTGCGACTGCA 1980  
Db 690 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 709  
Qy 1981 CACGTACATGGGGCATCTATAACAAAGGGGTGCTTACCTCAGCAGGCGAGGGAATA 2040  
Db 710 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 729  
Qy 2041 AAGAAACAAAGAGGAATTTCTAAGCCTATTAGAAGCCGTACATTTTACCAAAAAGCTAGCT 2100  
Db 730 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 749  
Qy 2101 ATTATACACTGCTCTGGACATCAGAAAGCTCAAGATCTCATATCCAGAGGAAACAGATG 2160  
Db 750 IleIleHisCysProGlyHisGlnLysAlaLysAspProIleSerArgGlyAsnGlnMet 769  
Qy 2161 GCTCAGCGGTGGCCAGGCGAGCCCGGGTGTAACTTCTCCCTATATAGAAATG 2220  
Db 770 AlaAspArgValAlaLysGlnAlaGlnGlyValAsnLeuLeuProMetIleGluThr 789  
Qy 2221 CCCAAAGCCCGAAGAACCCAGACAGTACACCTCTAGAAGACTGCGAAGATCAAAAAAG 2280  
Db 790 ProLysAlaProGluProGlyArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 809  
Qy 2281 ATAGACCACTGCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGGAATC 2340  
Db 810 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 829

QY	2341	CTGCCCCCAAGAAGGGTTAGAAATATGTCCTCAACAGATACATCGCTAAACCCACCTAGGA	2400
DB			
DB	830	LeuProHisLysLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly	849
QY	2401	ACTAAACACCTCGACAGCTGGTCAGAACATCCCCCTTATCATGTTCTGAGGCTACCAAGGA	2460
DB			
DB	850	ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuValArgLeuProGly	869
QY	2461	GTGGCTGACTCGGTGGTCAAACTTGTCGCCCTCCAGCTGGTTTAATCTTAATCTCTCC	2520
DB			
DB	870	ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer	889
QY	2521	AGAAATCGCTCCAGGGAACAGACTAAGGGGAAGCCACCCAGCGCTCTACTGGGAAGTCGAC	2580
DB			
DB	890	ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisIleTrpGluValAsp	909
QY	2581	TTCACTGAGGTAAAGCCGGCTAAATACGGAACAATACTTATTTGGTTTTTGTAGACACC	2640
DB			
DB	910	PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr	929
QY	2641	TTTTCAAGATCGGTAGAGGCTTTCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA	2700
DB			
DB	930	PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys	949
QY	2701	AAATACTGGAAGAAATTTTTCCAAGATTTCGAATACCTAAGGTAATAGGTCAGACAAT	2760
DB			
DB	950	LysIleLeuGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn	969
QY	2761	GGTCCAGCTTTTGTGGCCAGGTAAAGTCAGGAGCTGGCCAAAGATATTTGGGGATTGATGG	2820
DB			
DB	970	GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp	989
QY	2821	AAACTGCATTTGTGCATACAGACCCCAAGCTCAGACAGGTAGAGAGATGAATAGAAC	2880
DB			
DB	990	LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr	1009
QY	2881	ATTAAGAGACCCCTTACTAAATTCACCGCGAGACTGCGGTTAATGATTCGATAGCTCTC	2940
DB			
DB	1010	IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpMetAlaLeu	1029
QY	2941	CTGCCCTTTGTGCTTTTGTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCTAT	3000
DB			
DB	1030	LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr	1049
QY	3001	GAATTACTCTACGGGGGACCCCCCATTTGGTAGAAATGCTTCTGTGTACATAGTGTCTGAC	3060
DB			
DB	1050	GluLeuLeuTyrGlyGlyProProLeuAlaGluIleAlaPheAlaHisSerAlaAsp	1069
QY	3061	GTGCTGCTTTCCAGCCTTTGTTCTCTAGGCTCAAGGCACCTTGAGTCGGGTGAGACAACGA	3120
DB			
DB	1070	ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg	1089
QY	3121	CGCTGGAGGCAACTCCGGGAGGCCCTACTCAGGAGGAGGAGACTTGCAGATCCCAACATCGT	3180
DB			
DB	1090	AlaTrpLysGlnLeuArgGluAlaTyrSer--GlyGlyAspLeuGlnValProHisArg	1108
QY	3181	TTCCAAGTGGGAGATTCACTAGTCTAGTGTAGACCCACCGTCAGGAAACCTCGAGACTCGG	3240
DB			
DB	1109	PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg	1128
QY	3241	TGGAAGGGCCCTTATCTCGTACTTTTCACACACCAACGCGCTGGAAGTCGAAGGATC	3300
DB			
DB	1129	TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle	1148
QY	3301	TCCACCTGGATCCCATCCCGTAAACCCGCGCCACCTCCCGATTCCGGGGTGGAAA	3360
DB			
DB	1149	ProThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTrpLys	1168
QY	3361	CGCGAAAAGACTGAANAATCCCTTAAGCTTCGCCTCATCGGTGGTTCCTACTCTGTC	3420
DB			
DB	1169	AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal	1188

QY	3421	AATRACTCTCA	3432	
DB	1189	AsnAsnSer	1192	
RESULT	5			
Q901X4	90AMR			
ID	Q901X4_90AMR	PRELIMINARY;	PRT;	2378 AA.
AC	Q901X4			
DT	01-MAY-2000	(TrEMBLrel. 13, Created)		
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE	TYPE C proviral gag, pol and env genesand LTR (class A, clone			
DE	42).			
OS	Porcine endogenous retrovirus.			
OC	Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;			
OC	Gammaretrovirus; 1-Mammalian type C virus group.			
OX	NCBI_TaxID=61673;			
RP	[1]			
RN	NUCLEOTIDE SEQUENCE.			
RC	STRAUDN_F C;			
RA	Czauderna P., Fischer N., Boller K., Krach U., Kurth R., Toenjes R.R.;			
ET	"Molecular Characterization of Human-tropic and Replication-competent			
ET	Porcine Endogenous Retroviruses.";			
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.			
EMBL	AJ313817; CAB65340.1; -; Genomic_DNA.			
HSSP	P03385; IMOF.			
SMR	Q901X4; 7-98, 2227-2279.			
MEROPS	A02_020; -.			
GO	GO:0019028; C:viral capsid; IEA.			
GO	GO:0004190; F:aspartic-type endopeptidase activity; IEA.			
GO	GO:0003677; F:DNA binding; IEA.			
GO	GO:0004523; F:ribonuclease H activity; IEA.			
GO	GO:0003723; F:RNA binding; IEA.			
GO	GO:0003964; F:RNA-directed DNA polymerase activity; IEA.			
GO	GO:0005198; F:structural molecule activity; IEA.			
GO	GO:0006310; P:DNA recombination; IEA.			
GO	GO:0006508; P:proteolysis and peptidolysis; IEA.			
GO	GO:0006278; P:RNA-dependent DNA replication; IEA.			
GO	GO:0019068; P:viral assembly; IEA.			
InterPro	IPR002050; Env polyprotein; IEA.			
InterPro	IPR008981; FmU1Recept-bind.			
InterPro	IPR003036; Gag p30.			
InterPro	IPR000840; G_retro matrix.			
InterPro	IPR001995; Peptidase_A2 cat.			
InterPro	IPR000223; Peptidase_S26A.			
InterPro	IPR001969; Pept_Asp_AS.			
InterPro	IPR002156; RNaseH.			
InterPro	IPR001584; Rve.			
InterPro	IPR000477; RVTase.			
InterPro	IPR001878; Znf_CCHC.			
Pfam	PF01140; Gag_MA; 1.			
Pfam	PF02093; Gag_p30; 1.			
Pfam	PF00075; RNaseH; 1.			
Pfam	PF00665; rve; 1.			
Pfam	PF00077; RVP; 1.			
Pfam	PF00078; RVT; 1; 1.			
Pfam	PF00429; TLV_coat; 1.			
Pfam	PF00098; zf_CCHC; 1.			
SMART	SM00343; Znf_C2HC; 1.			
PROSITE	PS00141; ASP_PROTEASE; 1.			
PROSITE	PS00175; ASP_PROT_RETROV; 1.			
PROSITE	PS00879; RNAse_H; 1.			
PROSITE	PS00501; SPASE_I; 1.			
PROSITE	PS00158; ZF_CCHC; 1.			
CHAIN	1 524			
FT	CHAIN 525 1759			
FT	CHAIN 1719 2378			
SQ	SEQUENCE 2378 AA; 267306 MW; 6EC9DAEC1B4ED18A CRC64;			

Alignment Scores:

Pred. No.:	0	Length:	2378
Score:	5885.00	Matches:	1113
Percent Similarity:	98.7%	Conservative:	16



QY 2101 ATTATACACTGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATG 2160  
DB 1274 IleIleHisCysProGlyHisGlnLysAlaLysAepLeuIleSerArgGlyAsnGlnMet 1293  
QY 2161 GCTGACCGGGTTGCCAAGCAGCAGCCAGGGGTGTAACCTCTGCTTAAATGAAATG 2220  
DB 1294 AlaAepArgValAlaLysGlnAlaGlnAlaValAenLeuLeuProIleIleGluThr 1313  
QY 2221 CCCAAAGCCCAAGAACCCAGACGACAGTACACCCCTAGAGACTGGCAGAGATAAAAAG 2280  
DB 1314 ProLysAlaProGluProArgArgGlnTyrThrLeuGluAepTyrGlnGluIleLysLys 1333  
QY 2281 ATAGACCAAGTCTCTGACACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATC 2340  
DB 1334 IleAepGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle 1353  
QY 2341 CTGCCCAACAAGAGGGGTAGAATATGTCACAGATACATCTCTAAACCCACTAGGA 2400  
DB 1354 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 1373  
QY 2401 ACTAAACACCTGCAGCAGTCTGGTCAGAAACATCCCTTATCATGTTCTGAGGCTACCAAG 2460  
DB 1374 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 1393  
QY 2461 GTGGCTGACTCGTGGTCAAAATGTTGTGCTGCCAGCTGGTTAATGCTTAATCTCTCC 2520  
DB 1394 ValAlaAepSerValValLysHisCysValProCysGlnLeuValAenAlaAenProSer 1413  
QY 2521 AGAATGCTCCAGGAGAGACTAAGGGAGACCCAGCCGCTCACTGGGAAGTGAC 2580  
DB 1414 ArgIleProGlyLysArgLeuArgGlySerHisProGlyAlaHisTyrTrpGluValAep 1433  
QY 2581 TTCACTGAGTAAAGCCGCTAAATATACGAAACAAATACCTATTGTTTGTAGACACC 2640  
DB 1434 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 1453  
QY 2641 TTTTCAGATGGGTAGAGCTTATCTCTAAGAAAGAGACTCAACCGTGGTGGCTAAA 2700  
DB 1454 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 1473  
QY 2701 AAATACCTGGAGAAATTTTCCAGATTTGGATACCTAAGTAAATAGGTCAGACAT 2760  
DB 1474 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 1493  
QY 2761 GGTCAGCTTTGTTGCCAGTAAAGTCAGGACTGCGCAAGATATTGGGGATGATTGG 2820  
DB 1494 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspIrp 1513  
QY 2821 AAATGCTGTTGTGATACAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAAC 2880  
DB 1514 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 1533  
QY 2881 ATTAAGAGACCTTACTTAATTCACCGGAGACTGCGGTTAATGATTGATGATGCTCTC 2940  
DB 1534 IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAepTyrMetAlaLeu 1553  
QY 2941 CTGCCCTTTGTGCTTTTAGGGTTAGGAACACCCCTGCACAGTTTGCGCTGACCCCTAT 3000  
DB 1554 LeuProPheValLeuPheArgValArgAsnThrProGlnPheGlyLeuThrProTyr 1573  
QY 3001 GAATTAATCTACCGGGGACCCCCCAATTTGGTAGAAATTCCTCTGTACATAGTGTGAC 3060  
DB 1574 GluLeuLeuTyrGlyGlyProProProLeuValGluIleAlaSerValHisSerAlaAep 1593  
QY 3061 GTGCTGCTTTCCAGCCTTTGTTCTTAGGCTCAAGGCACCTTGAGTGGGTGAGAACGA 3120  
DB 1594 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTyrValArgGlnArg 1613  
QY 3121 GCGTGGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGAGACTTGCAGATCCACATCGT 3180  
DB 1614 AlaTrpLysGlnLeuArgGluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArg 1633  
QY 3181 TTCCAAGTGGAGATTCACTACGTTAGACGCCCGCTGACGAAACCTCGAGACTCGG 3240

DB 1634 PheGlnValGlyAepSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1653  
QY 3241 TGGAGGGCCCTTATCTCGTACTTTTGACACACCAACCGCTGTGAAAGTGAAGGAATC 3300  
DB 1654 TrpLysGlyProTyrLeuValLeuThrThrProThrAlaValLysValGluArgIle 1673  
QY 3301 TCCACCTGGATCCATGTCATCCACGTTAAACCGCGCCACCTCCCGATTCCGGGTGGA 3360  
DB 1674 ProThrTrpIleHisAlaSerHisVallyBProAlaProProAepSerGlyTrpLys 1693  
QY 3361 GCCGAAAGACTGAAATCCCTTAAGCTTCGCTCCATCGCTGGTTCCTTACTCTCTC 3420  
DB 1694 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1713  
QY 3421 AATAACCTCTCA 3432  
DB 1714 AsnAsnSerSer 1717  
RESULT 6  
Q8J4V6\_9GAMR  
ID Q8J4V6\_9GAMR PRELIMINARY; PRT; 1720 AA.  
AC Q8J4V6;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DS Gag-pol polypeptide.  
OS Porcine endogenous retrovirus B.  
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=194959;  
RN [1]  
RP NUCLEOTIDE SEQUENCE  
MEDLINE=22172947; PubMed=12185278;  
RA Bartosch B., Weiss R.A., Takeuchi Y.;  
RT "PCR-based cloning and immunocytological titration of infectious  
porcine endogenous retrovirus subgroup A and B";  
J. Gen. Virol. 83:2231-2240(2002).  
DR EMBL; AY093324; AAM29194.1; -; Genomic\_DNA.  
DR HSP; P03355; I16J.  
DR SMR; Q8J4V6; 7-98.  
DR MEROPS; A02.020; -.  
DR GO; GO:0019028; C:Viral capsid; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR003036; Gag\_p30.  
DR InterPro; IPR000840; G retro matrix.  
DR InterPro; IPR001995; Peptidase A2\_cat.  
DR InterPro; IPR001969; Pept\_Asp\_AS\_cat.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF01140; Gag\_MA\_1.  
DR Pfam; PF02093; Gag\_p30\_1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR Pfam; PF00098; zf\_CCHC; 1.  
DR SMART; SM00343; Znf\_C2HC; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS00879; RNASE\_H; 1.  
DR PROSITE; PS00158; ZF\_CCHC; 1.

KW Polyprotein.  
SQ SEQUENCE 1720 AA; 193393 MW; 7F9DCC2F5B834D8A CRC64;

Alignment Scores:  
Pred. No.: 0 Length: 1720  
Score: 5872.00 Matches: 1111  
Percent Similarity: 98.7% Conservative: 16  
Best Local Similarity: 97.3% Mismatches: 15  
Query Match: 95.0% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q8J4V6\_9GAMR (1-1720)

QY 1 ATGGGTGCGCAGGGCAACACGATCATCCATGGACTACCCGAGACAGATTGACTTGGGA 60  
DB MetGlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAspLeuGly 594

QY 61 GTGGACGGGTAAACCCACTCGTTCTGCTGATACCTGAGTGCACAGACCCCTCTTAGGT 120  
DB ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly 614

QY 121 AGAGACTTATTGACCAAGATGGAGCAAAATTTCTTTTGAACAAGGAAACCAAGAGTG 180  
DB ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal 634

QY 181 TCTGCAATAACAACCTTACTCTGTGTGACCTCCCAATTAGATGACGAATATCGACTA 240  
DB SerValAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspGluTyrArgLeu 654

QY 241 TACTCTCCCTAGTAAAGCCTGATCAAAATATACAATTTCTGTTGGAAACAGTTTCCCA 300  
DB TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGln 674

QY 301 GCCTGGGCGAACAACCGCAGGATGGTTTGGCAAGCAAGTTCCTCCCAACAAGTTATTCAA 360  
DB AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 694

QY 361 CTGAAGGCAGTGCACACGAGTGTCACTGACAGTACCCCTTGAGTAAAGAGCTCAA 420  
DB LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaArg 714

QY 421 GAAGGAATTCGGCGCATGTCCAAAGATTAACTCAAAGGCATCCTAGTTCTCTGTCCAA 480  
DB GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 734

QY 481 TCTCCCTGGAATATCTCCCTGCTACCGGTTAGAAGCCTGGGACTAATGACTATCGACCA 540  
DB SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 754

QY 541 GTACAGGACTTGAGAGGTCATTAACGGGTGCAGGATATACACCCACAGTCCCGAAC 600  
DB ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 774

QY 601 CTTTATAACCTCTTGTGTGCTCTCCACCCCAACGGAGCTGGTATACAGTATTGGACTTA 660  
DB ProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeu 794

QY 661 AAGGATGCTTCTTGTGCTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTTCGAA 720  
DB LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 814

QY 721 TGGAGAGATCAGGTACGGGAAGAACCGGCGAGTCACTGACCCGACTGCCCCACAGG 780  
DB TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 834

QY 781 TTCAGAACTCCCGACCATCTTTTACGAAGCCCTTACACAGAGACTGCCCAACTTCAGG 840  
DB PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 854

QY 841 ATCCAACACCTCTAGGTGACCTCTCTCAGTACGTGGATGACCTGCTTCTCGCGGAGCC 900  
DB IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla 874

QY 901 ACCAAACAGGACTGCTTTAGAAAGGCACGAAGGCACCTACTGCTGGAATTGTCTGACTAGGC 960  
DB ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 894

QY 961 TACAGAGCCTCTGTAAAGAGGCCAGATTTCGAGGAGAGAGGTAAACATACTTGGGGTAC 1020  
DB TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 914

QY 1021 AGTTTGGGGACGGCGAGCGATGGCTGACGGAGCAGCGAAGAAACCTGTAGTCCAGATA 1080  
DB SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 934

QY 1081 CCGGCCCAACACAGCAACAAATCAGAGAGTTTTTTGGGACAGCTGGGATTTGCGAGA 1140  
DB ProIleProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 954

QY 1141 CTGTGGATCCCGGGTTTTCGACCTTAGCAGCCCTTACTACCCGCTTAACCCAAAGAAAA 1200  
DB LeuTrpIleProGlyPheAlaThrLeuAlaProLeuTyrProLeuThrLysGluLys 974

QY 1201 GGGGAATTCCTCGGCTCCTGAGCAGCAGAGGCATTTGATGCTATCAAAAAGCCCTG 1260  
DB GlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 994

QY 1261 CTGAGCCACCTGCTCGGCCCTCCCTGACGTAAACCCCTTTTACCCCTTATGTGGAT 1320  
DB LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 1014

QY 1321 GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCCAAACCCCTTAGCACCATGGAGAGACCT 1380  
DB GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 1034

QY 1381 GTCCCTTACCTGTCAAAGAAGCTCGATCCTGTACCGAGTGGTTGGCCCATATGCTGAAG 1440  
DB ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLys 1054

QY 1441 GCTATCGCAGTGTGGCCATCTGCTCAAGGACCTGACAAATTTGACTTTGGGACAGAAAT 1500  
DB AlaIleAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 1074

QY 1501 ATAACTGTAATAGCCCCCATGCTTGGAGAACATCGTTGGGACGCCCCAGACCGATGG 1560  
DB IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 1094

QY 1561 ATGACCAACGCCCGCATGACCCACTATCAAGCCCTGCTTCTCAGAGAGAGGTTCACGTT 1620  
DB MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 1114

QY 1621 GCTCCACCGCGCTCTCAACCCCTGCCACTCTTCTGCTGAAGAGACTGTATGAACCCAGT 1680  
DB AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 1134

QY 1681 ACTCATGATTCGCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGAGACTTACAGAC 1740  
DB ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 1154

QY 1741 ATACCGCTGACTGAGAGAGTCTAACTGTTCTACTGACGGAAGCAGCTATGTGGTGAA 1800  
DB IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 1174

QY 1801 GGTAAGAGGATGGCTGGGGCGGTGGTGACGGGACCCCGACGATCTGGGCGCAGCAGC 1860  
DB GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpPheAspSer 1194

QY 1861 CTGCCGGAAGGAACCTTCAGCAAAAAGCTGAGTCTCATGGCCCTCAGCAAGCTTTGGCG 1920  
DB LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 1214

QY 1921 CTGGCCGAGGGAATTCATATAACATTTATACGACACAGGATGCTGCTTTGCCACTGCA 1980  
DB LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 1234

QY 1981 CACGTACATGGGGCCATCTATAAAACAAAGGGGTTGCTTACCTCAGCAGGGAGGAAATA 2040

Db 1235 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 1254  
 QY 2041 AAGAACAAAGAGAAATTTCTAAGCCTATTAGAAGCCGTACATTTTACCAAAAAGGCTAGCT 2100  
 Db 1255 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 1274  
 QY 2101 ATTATACATGCTCTCGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATG 2160  
 Db 1275 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 1294  
 QY 2161 GCTCACCGGTTCGCAAGCAGCCAGCCAGGGTCTTAACCTTCGCTTATATATGAATG 2220  
 Db 1295 AlaAspArgValAlaLysGlnAlaGlnAlaValAsnLeuLeuProIleIleGluThr 1314  
 QY 2221 CCCAAAGCCCCAGAACCCAGACGACAGTACACCCCTAGAAGACTGCGCAAGAGATATAAAAG 2280  
 Db 1315 ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 1334  
 QY 2281 ATAGACAGTTCTCTGAGACTCCGGAAGGGACCTGCTATACCTCAGATGGGAAGAAATC 2340  
 Db 1335 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle 1354  
 QY 2341 CTGCCCCACAAAGAGGTTAGATATGTCCTCAACAGATACATCTCTAAACCCACCTAGGA 2400  
 Db 1355 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 1374  
 QY 2401 ACTAAACACCTCGACAGCTGGTGCAGAACATCCCTTATCATGTTCTGAGCTACCAAGGA 2460  
 Db 1375 ThrLysHisLeuGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 1394  
 QY 2461 GTGCTGACTCGGTGGTCAACATGTTGTGCTCCGACGCTGGTTAATGCTTAATCTCTTC 2520  
 Db 1395 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 1414  
 QY 2521 AGATGCTCCAGGAAGACTAAGGGGAGCCACCCAGCGCTACTCGGAAGTGGAC 2580  
 Db 1415 ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 1434  
 QY 2581 TTCCTCAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGTTTGTGTAGACACC 2640  
 Db 1435 PheThrGluValLysProAlaLysTyrGlyAsnLysIleLeuLeuValPheValAspThr 1454  
 QY 2641 TTTTCAGATGGGTAGGCTTATCTCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 2700  
 Db 1455 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 1474  
 QY 2701 AAAATACTGAAGAAATTTTCCAGATTTGGATACCTAAGTAAATAGGCTCAGACAT 2760  
 Db 1475 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 1494  
 QY 2761 GGTCCAGCTTTTGTGGCCAGGTAAAGTCAGGACTGGCCCAAGATATTGGGGATTGATTGG 2820  
 Db 1495 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 1514  
 QY 2821 AAATGCTGATTTGTCATACAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAAC 2880  
 Db 1515 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 1534  
 QY 2881 ATTAAGAGACCTTACTTAATTTGACCGCGAGACTGGCGTTAATGTTGGATAGCTCTC 2940  
 Db 1535 IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeu 1554  
 QY 2941 CTGCTCTTCTGCTTTTGTAGGGTTAGGAACACCCCTGGACAGTTTGGCTGACCCCTAT 3000  
 Db 1555 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1574  
 QY 3001 GAATTAATCTACAGGGGACCCCCCTATTCCTGTTAGAAATTCCTCTGTACATAGTGTGAC 3060  
 Db 1575 GluLeuLeuTyrGlyGlyProProProProLeuValGluIleAlaSerValHisSerAlaAsp 1594  
 QY 3061 GTGCTGTTTCCAGGCTTTGTTCTCTAGGCTCAAGGCTTGAAGTGGTGGAGACACGA 3120

Db 1595 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1614  
 QY 3121 GCGTGGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGAGACTTGCAGATCCACATCGT 3180  
 Db 1615 AlaTrpLysGlnLeuArgGluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArg 1634  
 QY 3181 TTCAAAGTGGAGATTCACTTACGTTAGACGCCACCGTCAGGAAACCTCGAGACTCGG 3240  
 Db 1635 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1654  
 QY 3241 TGGNAGGCGCTTATCTCGTACTTTTGCACACCAACCGCTGTGAAAGTCAAGGAATC 3300  
 Db 1655 TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1674  
 QY 3301 TCACCTGGATCCATGTCATCCCGTTAAACCGCGCCACCTCCCGATTCGGGGTGGAAA 3360  
 Db 1675 SerThrTrpIleHisAlaSerHisValLysLeuAlaProProAspSerGlyTrpArg 1694  
 QY 3361 GCGGAAAGACTGAAATCCCTTAACTTCGCTCCATCGCGTTCCTTACTCTCTGTC 3420  
 Db 1695 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsn 1714  
 QY 3421 AATAAC 3426  
 Db 1715 AsnAsn 1716  
 RESULT 7  
 Q8J4V8\_9GAMR  
 ID Q8J4V8\_9GAMR PRELIMINARY; PRT; 1718 AA.  
 AC Q8J4V8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Porcine endogenous retrovirus A.  
 OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus;  
 OC 1-Mammalian type C virus group.  
 OX NCBI\_TaxID=194958;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 MDLINE=22172947; PubMed=12185278;  
 RA Bartosch B., Weiss R.A., Takeuchi Y.;  
 RT "PCR-based cloning and immunocytological titration of infectious  
 porcine endogenous retrovirus subgroup A and B";  
 J. Gen. Virol. 83:2231-2240(2002).  
 DR EMBL; AY099323; AAM29192.1; -; Genomic\_DNA.  
 DR HSSP; P03355; 116J.  
 DR SMR; Q8J4V8; 7-98.  
 DR MEROPS; A02.020; -.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO; GO:0019068; P:viral assembly; IEA.  
 DR InterPro; IPR003036; Gag\_p30.  
 DR InterPro; IPR000840; G retro matrix.  
 DR InterPro; IPR001995; Peptidase A2 cat.  
 DR InterPro; IPR001969; Pept\_Asp\_AS.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVase.  
 DR InterPro; IPR001878; Znf CCHC.  
 DR Pfam; PF01140; Gag\_MA\_1.  
 DR Pfam; PF02093; Gag\_p30; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.

DR Pfam: PF00078; RVT\_1; 1.  
DR Pfam: PF00098; ZF\_CCHC; 1.  
DR SMART: SM00343; ZNF\_C2HC; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS50175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS50879; RNASE\_H; 1.  
DR PROSITE; PS50158; ZF\_CCHC; 1.  
KW Polyprotein.  
SQ SEQUENCE 1718 AA; 193115 MW; 4E86215286C59246 CRC64;  
  
Alignment Scores:  
Pred. No.: 0 Length: 1718  
Score: 5866.50 Matches: 1111  
Percent Similarity: 98.4% Conservative: 15  
Best Local Similarity: 97.1% Mismatches: 17  
Query Match: 94.9% Indels: 1  
DB: 2 Gaps: 1  
  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q8J4V8\_9GAMR (1-1718)  
QY 1 ATGGGTGCCAGGGCAACACAGATCCATGGACTACCCGAAGACAGTTGACTTGGGA 60  
DB MetGlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAspLeuGly 594  
QY 61 GTGGACGGTAACCCACTCGTTCTGCTCATACCTGAGTGCACGACACCCCTCTTAGGT 120  
DB ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly 614  
QY 121 AGAGACTTATTGACCAAGATGGGACCAAAATTTCTTTTGAACAGGGAAACCCAGAAGTG 180  
DB ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal 634  
QY 181 TCTGCAATAACAACCTATCACTGTGTGTGACCTCCATTCAGATGAGCAATATCGACTA 240  
DB SerValAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu 654  
QY 241 TACTCTCCCTAGTAAAGCCTGATCAAAATATACAAATTCGTGTTGGACAGTTTCCCCAA 300  
DB TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGlnPheProGln 674  
QY 301 GCCTGGGAGAAACCGCAGGGATGGGTTTGGCAAGCAAGTTCCCCCAACAAGTTATTCAA 360  
DB AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 694  
QY 361 CTGAAGGCCAGTGCCACACAGTGTCACTCAGACAGTACCCCTGAGTAAAGACTCAA 420  
DB LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaArg 714  
QY 421 GAAGGAATTCGGCCGATGTCCAAAGATTAAATCAACAGGCGATCCTAGTTCTGTGCCAA 480  
DB GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 734  
QY 481 TCTCCTCGGAATACTCCCTGCTACCGGTTAGAAAGCCTGGGACTTAATGACTATCGACCA 540  
DB SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 754  
QY 541 GTACAGGACTTGAGAGGTCATAAACGGGTGCAGGATATACACCCCAACAGTCCCGAAC 600  
DB ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 774  
QY 601 CCTTATAACCTCTGTGTGCTCCACCCACCAACGGAGCTGGTATACAGTATTGGACTTA 660  
DB ProTyrAsnLeuLeuSerAlaLeuProGlnArgAsnTrpTyrThrValLeuAspLeu 794  
QY 661 AAGGATGCTCTCTTCTGCTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTTCGAA 720  
DB LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 814  
QY 721 TGGAGAGATTCAGGTACGGGAAGAACCGGGCGAGTCACTGGAACCCGACTGCCCAAGGG 780  
DB TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 834  
QY 781 TTCAAGAACTCCCGACCATCTTTTGACGAAGCCCTTACACAGAGACCTTGGCCAACTTCAGG 840

DB PhelysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 854  
QY 841 ATCCAAACACCTCTAGGTGACCTCTCTCAGTACCTGATGATGACCTGCTTCTGGCGGAGCC 900  
DB IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla 874  
QY 901 ACCAAACAGGACTCTGTAGAGGACCAAGGACACTACTGCTGGAAATTGTCTGACCTAGGC 960  
DB ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuGluLeuSerAspLeuGly 894  
QY 961 TACAGAGCTCTGCTAAGAAGGCCCCAGATTTCAGGAGAGAGGTAACATACTTTGGGGTAC 1020  
DB TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 914  
QY 1021 AGTTTGGGGACGGCGACGATGCTGACGAGGACGAGGAAGAAACACTGTAGTCCAGATA 1080  
DB SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 934  
QY 1081 CCGGCCCAACACAGACCAACAAATGAGAGATTGTTGGGACAGCTGGATTGTCAGA 1140  
DB ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 954  
QY 1141 CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCTACTTACCCGCTAACCAAGAAAAA 1200  
DB LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLys 974  
QY 1201 GGGAAATCTCTGGGCTCTGAGCACAGAAAGCATTTGATGTATCAAAAAGGCCCTG 1260  
DB GlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 994  
QY 1261 CTGAGCCACCTGCTGCGCCCTCCCTGAGCTAACTAAACCTTTTACCTTTATGTGAT 1320  
DB LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 1014  
QY 1321 GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTTAGGACCATGAGAGACCT 1380  
DB GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgTrp 1034  
QY 1381 GTCCCTACCTGTCAAAGAGCTCGATCTGTAGCAGTGTGTGGCCCATATGCTGAAG 1440  
DB ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTyrProValCysLeuLys 1054  
QY 1441 GCTATCCAGCTGGCCCATACTCGTCAAGGACCTGACAAATGACTTTTCGGACAGAAAT 1500  
DB AlaIleAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 1074  
QY 1501 ATAACTGTAAATAGCCCCCATGTCATTGGAGAACATCGTTTCGGCAGCCCCCAGACCGATG 1560  
DB IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 1094  
QY 1561 ATGACCAAGCCCGCATGACCCATCATCAAAAGCCTGTCTTCACAGAGAGGTCACGTTTC 1620  
DB MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 1114  
QY 1621 GCTCCACGACCGCTCAACCTGCGCACTCTCTGCTGAGAGACTGATGAACCACTG 1680  
DB AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 1134  
QY 1681 ACTCATGATTCCTCATCACTATTGATTGAGGAGACTTGGGGTCCGCAAGGACCTTACAGAC 1740  
DB ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 1154  
QY 1741 ATACCGTGTGAGGAGAGTCTAACTGTGTTCTACTGACGGAAGCAGACTATGTGGTGAA 1800  
DB IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 1174  
QY 1801 GGTAAAGGATGGCTGGGGCGGCTGTGTGACGCGACCCGACGATCTGGCGCAGCAGC 1860  
DB GlyLysArgMetAlaGlyAlaValValAspGlyThrArgThrIleTrpAlaSerSer 1194  
QY 1861 CTGCCGGAAGGAACCTTCAGCAAAAAAGGCTCATGCTCATGGCCCTCACCAAGCTTTGCGG 1920



Db 1195 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 1214  
Qy 1921 CTGCCCGAAGGAAATCATTAACATTTATACGACAGCAGGTATGCCCTTTCGCACTGCA 1980  
Db 1215 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 1234  
Qy 1981 CACCTACATGGGGCCTATATAACAAGGGGGTGTCTTACTCAGCGAGGAGGAAATA 2040  
Db 1235 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 1254  
Qy 2041 AAGAACAAAGAGAAATCTTAAGCCTATTAGAAGCGGTACATTTTACCAAAAAGGCTAGCT 2100  
Db 1255 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 1274  
Qy 2101 ATTATACATGTCTCGACATPCAGAAAGCTTAAGATCTCATATCCAGAGAAACACAGATG 2160  
Db 1275 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 1294  
Qy 2161 GCTCACCGGTTGCCAAGCAGCGCCAGGGGTGTAACTTCTGCCTATAATAGAATG 2220  
Db 1295 AlaAspArgValAlaLysGlnAlaGlnAlaValAsnLeuLeuProIleIleGluThr 1314  
Qy 2221 CCCAAAGCCCGACAGACAGCAGTACACCTTAGAAGACTGCGAAGAGATAAAAAG 2280  
Db 1315 ProLysAlaProGluProArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 1334  
Qy 2281 ATAGACCATGTTCTCGAGACTCCGGAAGGAGCTTACTATACCTCAGATGGGAAGAAATC 2340  
Db 1335 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle 1354  
Qy 2341 CTGCCCCAAGAAAGGGTTAGATATGTCCAACAGATACATCGTCTAAACCCACTAGGA 2400  
Db 1355 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 1374  
Qy 2401 ACTAAACACCTCGACGAGTGGTCAGAACATCCCTTATCATGTTCTGAGCTACACGGA 2460  
Db 1375 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 1394  
Qy 2461 GTGCTGACTCGGTGGTCAACACTGTGTGCTCCGCCAGCTGGTTAATGCTAATCCTTCC 2520  
Db 1395 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 1414  
Qy 2521 AGAATGCTCCAGGAAAGAGACTAAGGGGAAGCCACCCAGCGCTCACTGGGAAGTGGAC 2580  
Db 1415 ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 1434  
Qy 2581 TTCCTGAGCTAAGCCGGCTAAATACGGAACCAATACCTATTGTTTGTAGACACC 2640  
Db 1435 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 1454  
Qy 2641 TTTTTCAGGATGGGTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 2700  
Db 1455 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValAlaLys 1474  
Qy 2701 AAAATCTGGAAGAAATTTTCCAAGATTGTGAATACCTAAGGTAAATAGGGTCAACAAT 2760  
Db 1475 LysIleLeuGluGluIlePheProArgPheGlyLeuProLysValIleGlySerAspAsn 1494  
Qy 2761 GSTCCAGCTTTCTGCCCAGGTAAAGTACGAGGACTGCGCAAGATATGGGGATGATGG 2820  
Db 1495 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 1514  
Qy 2821 AAATCTGATTTGTGCATACAGACCCAAAGCTCAGGACAGGTAGAGAGATGAATAGAAC 2880  
Db 1515 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 1534  
Qy 2881 ATTAAGAGACCTTACTAAATTCAGCGGAGACTGGCGTTAATGATTGATGAGTCTC 2940  
Db 1535 IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpMetAlaLeu 1554  
Qy 2941 CTGCCCCTTTGTGCTTTTAGGGTTAGGAACACCCCTGGACAGTTGGCTGACCCCTAT 3000  
Db 1555 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1574

Qy 3001 GAATTACTTACGGGGGACCCCCCATTTGGTAGAAATTGCTTCTGTACATAGTGTGAC 3060  
Db 1575 GluLeuLeuTyrGlyGlyProProLeuAlaGluIleAlaPheAlaHisSerAlaAsp 1594  
Qy 3061 GTGCTGCTTTCCAGCAGCTTTGTTCTTAGGCTCAAGGCATTGAGTGGGTGAGACACGA 3120  
Db 1595 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1614  
Qy 3121 GCGTGGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGAGACTTGCAGATCCCATCGT 3180  
Db 1615 AlaTrpLysGlnLeuArgGluAlaTyrSer---GlyGlyAspLeuGlnValProHisArg 1633  
Qy 3181 TTCCAAGTGGAGATTCACTTACGTTAGACGCCACCGTGCAGGAAACCTCGAGACTCGG 3240  
Db 1634 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1653  
Qy 3241 TGGAGGGCCCTTATCTCGTACTTTTGACACACCAACCGGTGTGAAGTGAAGGAATC 3300  
Db 1654 TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1673  
Qy 3301 TCCACCTGGATTCATGCATCCCGTAAACCGGCGCCACTCCCGATTGGGGTGGAAA 3360  
Db 1674 ProThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTrpLys 1693  
Qy 3361 GCCGAAAGACTGAAAATCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTC 3420  
Db 1694 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1713  
Qy 3421 AATAACCTCTCA 3432  
Db 1714 AsnAsnSerSer 1717  
RESULT 8  
Q9Q1X5\_9GAMR  
ID Q9Q1X5\_9GAMR PRELIMINARY; PRT; 2376 AA.  
AC Q9Q1X5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Type C proviral gag, pol and env genes and LTR (class B, clone 33).  
DE OS Porcine endogenous retrovirus.  
DE OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
DE OC Gamma retroviruses; 1-Mammalian type C virus group.  
DE OX NCBI\_TaxID=61673;  
DE RN [1]  
DE RP NUCLEOTIDE SEQUENCE.  
DE RC STRAIN=type C;  
DE RX MEDLINE=20219394; PubMed=10756014;  
RX DOI=10.1128/JVI.74.9.4028-4038.2000;  
RA Czauderna F., Fischer N., Boller K., Kurth R., Toenjes R.R.;  
RT "Establishment and characterization of molecular clones of porcine endogenous retroviruses replicating on human cells";  
RL J. Virol. 74:4028-4038(2000).  
DR EMBL; AJ133816; CAB65339.1; -; Genomic\_DNA.  
DR HSSP; P03385; 1MOP.  
DR SNR; Q9Q1X5; 7-98, 472-518, 2225-2277.  
DR MEROPS; A02.020; -.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR002050; Env\_polyprotein.  
DR InterPro; IPR008981; FmULVrecept-bind.  
DR InterPro; IPR003036; Gag\_p30.



DR	InterPro; IPR000840; G_retro_matrix.	Db	754	ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn	773
DR	InterPro; IPR001995; Peptidase_A2_cat.	QY	601	CCTTATAACCTCTTGTGTGCTCTCCACCCCAACGAGCTGGTATACAGTATTGGACTTA	660
DR	InterPro; IPR001969; Pept_Asp_AS.	QY	601		
DR	InterPro; IPR002156; RNaseH.	Db	774	ProTyrAsnLeuLeuSerAlaLeuProGluArgAsnTrpTyrThrValLeuAspLeu	793
DR	InterPro; IPR001584; Rve.	Db	774		
DR	InterPro; IPR000477; RVise.	QY	661	AAGGATGCCTTCTTGTGCTGAGATTACACCCCACTAGCCAAACCACTTTTGGCTTCGAA	720
DR	InterPro; IPR001878; Znf_CCHC.	QY	661		
DR	Pfam; PF01140; Gag_MA; 1.	Db	794	LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu	813
DR	Pfam; PF02093; Gag_p30; 1.	Db	794		
DR	Pfam; PF00075; RNaseH; 1.	QY	721	TGGAGAGATCCAGGTACGGGAAGAACCGGCAGCTCACCTGGACCCGACTGCCCAAGGG	780
DR	Pfam; PF00665; rve; 1.	QY	721		
DR	Pfam; PF00077; RVP; 1.	Db	814	TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly	833
DR	Pfam; PF00078; RVT_1; 1.	Db	814		
DR	Pfam; PF00429; TLV_coat; 1.	QY	781	TTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGG	840
DR	Pfam; PF00098; zf-CCHC; 1.	Db	834	PhelysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg	853
DR	SMART; SM00343; Znf_C2HC; 1.	Db	834		
DR	PROSITE; PS00141; ASP_PROTEASE; 1.	QY	841	ATCCAACACCTCAGGTGACCTCTCCAGTACGTGATGACCTGCTTCTGGCGGAGCC	900
DR	PROSITE; PS0175; ASP_PROT_RETROV; 1.	QY	841		
DR	PROSITE; PS50879; RNASE_H; 1.	QY	841		
DR	PROSITE; PS50158; ZF_CCHC; 1.	Db	854	IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla	873
FT	CHAIN 1 524 Gag protein.	Db	854		
FT	CHAIN 525 1761 Pol protein.	QY	901	ACCAACACGAGCTGTAGNAGGCACGAAGGCACACTACTGCTGGAATTGTCTGACTAGGC	960
FT	CHAIN 1720 2376 Env protein.	QY	901		
SQ	SEQUENCE 2376 AA; 266625 MW; 37DD39D3099DA4AF CRC64;	Db	874	ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly	893
Alignment Scores:					
Pred. No.:	0	Length:	2376		
Score:	5864.00	Matches:	1109		
Percent Similarity:	98.5%	Conservative:	16		
Best Local Similarity:	97.1%	Mismatches:	17		
Query Match:	94.8%	Indels:	0		
DB:	2	Gaps:	0		
US-10-723-552-3_COPY_2307_5741 (1-3435) x Q90IX5_9GAMR (1-2376)					
QY	1	ATGGGTGCCACAGGCGCAACACAGTATCCATGGACTACCCGAGCAACAGTTGACTTGGGA	60		
Db	574	MetGlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAspLeuAla	593		
QY	61	GTGGAGCGGTAACCACTGTTTCTGGTCATACCTGAGTCCGACGACCCCTCTTAGGT	120		
Db	594	ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly	613		
QY	121	AGAGACTTATTGACCAAGATGGGAGCAAAATTTCTTTGAACAGGAAACACAGAGTG	180		
Db	614	ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal	633		
QY	181	TCTGCAATAACAAACCTTATCAGTGTGTGACCTCCCAATTAGATGACGAATATCGACTA	240		
Db	634	SerValAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu	653		
QY	241	TACTCTCCCTAGTAAGCCTGATCAAAATATACAATTTCTGGTTGGACAGTTTCCCAA	300		
Db	654	TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGln	673		
QY	301	GCCTGGGCGAAGAACCGCAGGATGGTTTGGCAAGCAAGTTCCCCCAAGTTATTCAA	360		
Db	674	AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln	693		
QY	361	CTGAAGGCCAGTGCCACACCACTGTGCTGACGACAGTACCCCTTGAGTAAAGAAAGCTCAA	420		
Db	694	LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaArg	713		
QY	421	GAAAGAAATTCGGCCGATGTCCAAAGATTAAATCCAACAGGGCATCTTAGTTCTCGTCAA	480		
Db	714	GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln	733		
QY	481	TCTCCCTGGATATCTCCCTGCTACCGTTAGAAAGCTCGGACTAATGACTATCGACCA	540		
Db	734	SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro	753		
QY	541	GTACAGGACTTGAAGAGGTCAATAAAGCGGTGCAGGATATACACCAACAGTCCCGAAC	600		
Db	1114	AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluThrAspGluProVal	1133		

QY 1691 ACTCATGATGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGACCTTACAGAC 1740  
Db ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 1153  
QY 1741 ATACCGCTGACTGGAGAGTGTCTAACTGGTTCCTACTGACGGAAGACGATATGTGGTGA 1800  
Db IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 1173  
QY 1801 GGTAAAGAGATGGCTGGGGCGGGTGTGACGCGGACCCACGACATCTGGGCGCAGCAGC 1860  
Db GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrHisThrIleTrpAlaSerSer 1193  
QY 1861 CTGCGGGAAGAACTTTCAGCAAAAGGCTGAGCTCATGGCCCTCACCAAGCTTTGGCG 1920  
Db LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 1213  
QY 1921 CTGCGGGAAGGAATCATAAACATTTATAGGACAGCAGGATGCTTTGGCAGCTGCA 1980  
Db LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 1233  
QY 1981 CACGTACATGGGGCCATCTATAAACAAAGGGGTTGCTTACCTCAGCAGGAGGAGAAATA 2040  
Db HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 1253  
QY 2041 AAGAACAAAGAGGAAATTTCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCT 2100  
Db LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 1273  
QY 2101 ATTATACATCTCTCGNACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATG 2160  
Db IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 1293  
QY 2161 GCTCACCGGGTTCGCAAGCAGCAGCCAGGGTGTAACTTCGCTCTAATAAGAAATG 2220  
Db AlaAspArgValAlaLysGlnAlaGlnAlaValAsnLeuLeuProIleIleGluThr 1313  
QY 2221 CCCAAAGCCCAAGAACCCAGACGACGACAGTACACCCCTAGAAAGACTGGCAAGAGATAAAAAAG 2280  
Db ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 1333  
QY 2281 ATAGACAGATCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATC 2340  
Db IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle 1353  
QY 2341 CTGCCCCACAAAGAGGTTAGATATGTCACACAGATACATCTCTAAACCCACTAGGA 2400  
Db LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 1373  
QY 2401 ACTAAACACCTGCAGCAGTGTGGTCAGAACATCCCTTATCATGTTCTCGAGGCTACCAAGGA 2460  
Db ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuValArgLeuProGly 1393  
QY 2461 GTGGCTGACTCGGTGGTCAACATGTGTGTCCTCCGACGCTGGTAAATGCTTAATCTCTCC 2520  
Db ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 1413  
QY 2521 AGAATGCTCCAGGGAAGACTAAGGGAGCCACCCAGCGCTCACTGGGAGTGGAC 2580  
Db ArgIleProProGlyLysArgLeuArgLysSerHisProGlyAlaHisTrpGluValAsp 1433  
QY 2581 TTCACTGAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGTTTGTGTAGACACC 2640  
Db PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 1453  
QY 2641 TTTTCAGATGGGTAGAGGCTTATCTCTACTAAGAAAGAGACTTCAACCGGTGGTGGTAAA 2700  
Db PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 1473  
QY 2701 AAAATACGAGGAAGAAATTTTCCAGATTTGGATACCTAAGTAAATAGGGTCAGACAAT 2760  
Db LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 1493

QY 2761 GGTCCAGCTTTTGTGGCCAGGTAAAGTCCAGGACTGGCCAAAGATATTGGGATTCATTGG 2820  
Db GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyLysAspTrp 1513  
QY 2821 AAACCTGATTTGTGATCAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAAC 2880  
Db LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 1533  
QY 2881 ATTTAAAGAGACCTTTACTTAAATTGACCGCGAGAGACTGCGTTAATGATTGGATGCTCTC 2940  
Db IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeu 1553  
QY 2941 CTGCCCCCTTTGTGCTTTTGTAGGTTAGGAACACCCCTGACAGAGTTTGGGCTCACCCTCAT 3000  
Db LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1573  
QY 3001 GAATTACTCTACGGGGGACCCCTTCTAGGCTCAAGGACCTTGAGTGGGTGAGACAACGA 3060  
Db GluLeuLeuTyrGlyGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1593  
QY 3061 GTGCTGCTTTCCCAAGCCTTTGTTCTTAGGCTCAAGGACCTTGAGTGGGTGAGACAACGA 3120  
Db ValLeuLeuSerGlnProLeuPheSerArgLysLysAlaLeuGluTrpValArgGlnArg 1613  
QY 3121 GCGTGGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGGAGACTTGCAGATCCCATCGT 3180  
Db AlaTrpLysGlnLeuArgGluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArg 1633  
QY 3181 TTCCAAAGTGGAGATTTCAGTCTAGTTAGACGCCACCGTCGAGGAAACCTTCGAGACTCGG 3240  
Db PheGlnValGlyAspSerValTyrValArgHisArgAlaGlyAsnLeuGluThrArg 1653  
QY 3241 TGGAAAGGCGCTTATCTCGTACTTTTCACACACCAACGCGCTGTAAGTTCGAAGGAATC 3300  
Db TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1673  
QY 3301 TCACCTCGATTCATGTCATCCACGTTTAAACCCGCGCCACCTCCCGATTTCGGGTGGA 3360  
Db SerThrTrpIleHisThrSerHisValLysLeuAlaProProAspSerGlyTrpArg 1693  
QY 3361 GCGGAAAGACTGAAATTCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTC 3420  
Db AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsn 1713  
QY 3421 AATAAC 3426  
Db 1714 AsnAsn 1715  
RESULT 9  
Q8UMP5\_9GAMR  
ID Q8UMP5\_9GAMR PRELIMINARY; PRT; 1146 AA.  
AC Q8UMP5;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Pol.  
OS Porcine endogenous retrovirus.  
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21568379; PubMed=11711616;  
RX DOI=10.1128/JVI.75.24.12252-12265.2001;  
RA Herring C., Quinn G., Bower R., Parsons N., Logan N.A., Brawley A.,  
RA Elsom K., Whittam A., Fernandez-Suarez X.M., Cunningham D.,  
RA Onions D., Langford G., Scobie I.,  
RT "Mapping full-length porcine endogenous retroviruses in a large white  
pig".  
RL J. Virol. 75:12252-12265(2001).  
DR EMBL; AY056035; AAL38193.1; -; Genomic\_DNA.  
DR HSP; P03355; 116J.  
DR MBROPS; A02.020; -.

DR GO: 0004190; F: aspartic-type endopeptidase activity; IEA.  
DR GO: 0003677; F: DNA binding; IEA.  
DR GO: 0004523; F: ribonuclease H activity; IEA.  
DR GO: 0003723; F: RNA binding; IEA.  
DR GO: 0003964; F: RNA-directed DNA polymerase activity; IEA.  
DR GO: 0006310; P: DNA recombination; IEA.  
DR GO: 0006508; P: proteolysis and peptidolysis; IEA.  
DR GO: 0006278; P: RNA-dependent DNA replication; IEA.  
DR InterPro: IPR001995; Peptidase\_A2\_cat.  
DR InterPro: IPR002156; RNaseH.  
DR InterPro: IPR001584; Rve.  
DR InterPro: IPR000477; RVTse.  
DR Pfam: PF00075; RNaseH; 1.  
DR Pfam: PF00665; rve; 1.  
DR Pfam: PF00077; RVP; 1.  
DR Pfam: PF00078; RVT\_1; 1.  
DR PROSITE: PS0175; ASP\_PROT\_RETROV; 1.  
DR PROSITE: PS0879; RNASE H; 1.  
SQ SEQUENCE 1146 AA; 128417 MW; 1244B29FC5C270DE CRC64;

## Alignment Scores:

Pred. No.:	0	Length:	1146
Score:	5863.00	Matches:	1109
Percent Similarity:	98.6%	Conservative:	17
Best Local Similarity:	97.1%	Mismatches:	16
Query Match:	94.8%	Indels:	0
DB:	2	Gaps:	0

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q8UMP5\_9GAMR (1-1146)

QY	1	ATGGTGCCACGAGGCAACAACAGTATCCATGGACTACCCGGAAGAACAGTTGACTTGGGA	60
Db	1	MetGlyAlaThrGlyGlnArgGlnIyrProTrpThrThrArgArgThrValAspLeuGly	20
QY	61	GTGGACGGGTAAACCACCTCGTTTCTGGTCATACCTGAGTGGCCAGCACCCCTCTTAGGT	120
Db	21	ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly	40
QY	121	AGAGACTTATTGACCAAGATGGAGCAACAATTTCTTTGAACAAGGAAACCAAGAGTG	180
Db	41	ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal	60
QY	181	TCTGCAATAACAACCTATCACTGTGTGACCTCCAAATTAGATGACGAATATCGACTA	240
Db	61	SerValAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu	80
QY	241	TACTCTCCCTAGTAAAGCCTGATCAAAATATACAATTCCTGGTTGGACAGTTTCCCCAA	300
Db	81	TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGln	100
QY	301	GCCTGGGAGAAACCGGAGGATGGTTTGGCAAGCAAGTTCCCCCAAGATTATTCAA	360
Db	101	AlaTrpAlaGluThrAlaGlyIleGlyLeuAlaLysGlnValProProGlnValIleGln	120
QY	361	CTGAAGGCCAGTGCCACACCAAGTGTCACTGACAGACTACCCCTTGAGTAAAGAAGCTCAA	420
Db	121	LeuLysAlaSerAlaThrProValSerValArgGlnIyrProLeuSerArgGluAlaArg	140
QY	421	GAAGGAATTCGGCGCATGTCCAAGATTAAATCCAACAGGGCATCTAGTTCTCTGTCAA	480
Db	141	GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnIleLeuAlaProValGln	160
QY	481	TCTCCTGGAAATACCTCCCTGCTACCGTTAGAAAGCCTGGGACTAATGACTATCGACCA	540
Db	161	SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro	180
QY	541	GTACAGGACTTGAAGAGTCAATAAAGGGTGCAGGATATACCCCAACAGTCCCGAAC	600
Db	181	ValGlnAspLeuAsgluValAsnLysArgValGlnAspIleHisProThrValProAsn	200
QY	601	CCTTATAACCTCTTGTGTCTCTCCACCCCAACGAGCTGGTATACAGTATTGGACTTA	660
Db	201	ProTyrAsnLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeu	220

QY	661	AAGGATCCCTTCTTCTGCTGAGATTACCCCACTAGCCAAACCACTTTTTCCTTCGAA	720
Db	221	LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu	240
QY	721	TGGAGAGATCCAGGTACGGGAAGAACCGGCAGCTCACCTGGACCCGACTGCCCAAGGG	780
Db	241	TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly	260
QY	781	TTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGG	840
Db	261	PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg	280
QY	841	ATCCAACACCTCAGGTGACCTCCTCAGTACGTGATGACCTGCTTCTGGCGGAGGCC	900
Db	281	IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla	300
QY	901	ACCAACAGGACTGCTTTAGAGGCACGAAGCACTACTGCTGGAATTGTCTGACCTAGGC	960
Db	301	ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuGluLeuSerAspLeuGly	320
QY	961	TACAGACCTCTGCTAAGAAGCCACATTTGACGAGAGAGGTAAACATCTTGGGTAC	1020
Db	321	TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr	340
QY	1021	AGTTTGGGACGGCGAGCGATGGCTGACGAGGCACGGAAGAACTGTAGTCCAGATA	1080
Db	341	SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysThrValValGlnIle	360
QY	1081	CCGGCCCCAACCAAGCAAAATGAGAGAGTTTTTGGGGACAGCTGGATTTTGAGA	1140
Db	361	ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg	380
QY	1141	CTGTGATCCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTAACCAAGAAAAA	1200
Db	381	LeuTrpIleProGlyPheAlaThrLeuAlaProLeuTyrProLeuThrLysGluLys	400
QY	1201	GGGGAATTCCTGGGCTCCTGACACACAGAGCATTTGATGTATCAAAAGGCCCTG	1260
Db	401	GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu	420
QY	1261	CTGAGCCACCTGCTGCGCCCTCCCTGACGTAACTAAACCCCTTTACCTTTATGTGAT	1320
Db	421	LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp	440
QY	1321	GAGGTAAGGAGTAGCCCGGGAGTTTAAACCAACCCCTAGCACCATTGAGAGAGACCT	1380
Db	441	GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgPro	460
QY	1381	GTGSCCTACCTGTCAAGAAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCTGAAG	1440
Db	461	ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLys	480
QY	1441	GCTATCAGCTGTGGCCATACCTGGTCAAGGACCTGACAAATTCACCTTTGGGACAGAA	1500
Db	481	AlaIleAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuArgGlnAsn	500
QY	1501	ATAACTGTATAGCCCCCATGCTGAGAGACATCGTTGGGAGCCGCCAGACCGATGG	1560
Db	501	IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp	520
QY	1561	ATGACCAACGCCCGCATGCCACCTATCAAGCCCTGCTTCTCACAGAGAGGCTCACGTTT	1620
Db	521	MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe	540
QY	1621	GCTCCACGACCGCTCTCAACCCCTGCCACTCTTCTGCTGAAGAGACTGTAAACCAAGT	1680
Db	541	AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal	560
QY	1681	ACTCATGATTCCTCAACTATTGATTGAGGAGACTGGGTCCCGAAGGACCTTACAGAC	1740
Db	561	ThrHisAspCysHisGlnLeuLeuIleGluThrGlyValArgLysAspLeuThrAsp	580

QY 1741 ATACCGCTGACTGGAGAGTCTTAACCTGGTTCCTACTGACGGAAGCAGCTATGTGGTGA 1800  
DB 581 ILeProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 600  
QY 1801 GGTAAAGAGATGGCTGGGGCGCGTGTGTGACCGGACCCCGACGATCTGGGCCAGCAGC 1860  
DB 601 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 620  
QY 1861 CTGCGCGAAGAACTTTCAGCAAAAGGCTGAGCTCATGGCCCTCAGCAAGCTTTGCGG 1920  
DB 621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
QY 1921 CTGCGCGAAGGAAATCCATAAATTCATTAAGGACGAGCAGGATGCTTTGCCAGCTGCA 1980  
DB 641 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 660  
QY 1981 CACCTACATGGCGGCATCTATAAACAAGGGGTTGCTTACCTCAGCAGGAGGAGAAATA 2040  
DB 661 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
QY 2041 AAGAACAAAGAGGAAATTTCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCT 2100  
DB 681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 700  
QY 2101 ATTATACACTGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATG 2160  
DB 701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720  
QY 2161 GCTCACCGGGTTGCAAGCAGCAGCCAGGGTGTAAACCTTCCTGCTTAATAGAAATG 2220  
DB 721 AlaAspArgValAlaLysGlnAlaGlnAlaValAsnLeuLeuProIleIleGluThr 740  
QY 2221 CCCAAAGCCCCAGAACCCAGACGACAGTACACCCCTAGAAAGCTGGCAAGAGATAAAAAG 2280  
DB 741 ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760  
QY 2281 ATAGACACTGCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATC 2340  
DB 761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle 780  
QY 2341 CTGCCCCACAAGAAAGGTTAGATATGTCACAGATACATCGCTAAACCCACTAGGA 2400  
DB 781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
QY 2401 ACTAAACACCTGCAGCAGTGGTGCAGAACATCCCTTATCATGTTCTGAGGCTACCAAGGA 2460  
DB 801 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
QY 2461 GTGGCTGACTCGGTGGTCAACATTTGTGTGCTGCCCTGCCAGCTGGTTAATGCTAATCTTCC 2520  
DB 821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
QY 2521 AGAATGCTCCAGGGAAGACTAAAGGGGAAGCCACCGCGCTCACTGGGAGTGCAC 2580  
DB 841 ArgIleProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860  
QY 2581 TTCACTCAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGTTTGTAGACACC 2640  
DB 861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
QY 2641 TTTTCAGATGGGTAGAGCTTATTCCTACTAAGAAAGACTTCAACCGTGGTGGCTTAA 2700  
DB 881 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900  
QY 2701 AAAATCTGGAAGAAATTTTCCAGATTTGGATACCTAAGCTTAATAGGTCAGCAAT 2760  
DB 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
QY 2761 GGTCCAGCTTTTGTGCCAGGTAAAGTCAGGAGCTGGCCCAAGATATTGGGGATTGATTGG 2820  
DB 921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
QY 2821 AAATGCAATTGTGCATACAGACCCCAAGCTCAGGACAGGTAGAGAGGATGAATAGAAC 2880

DB 941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
QY 2881 ATTAAGAGACCCCTTACTTAAATTCACCGCGAGACTGCGGTTAATGATTGGATAGCTCTC 2940  
DB 961 IleLysGluThrLeuThrLysLeuThrGluThrGlyIleAsnAspTrpIleAlaLeu 980  
QY 2941 CTGCCCCCTTTGCTGCTTTTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCTAT 3000  
DB 981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
QY 3001 GAATTACTCTACGGGGACCCCGCCCATTTGGTAGAAATTCCTTCTGTACATAGTCTGAC 3060  
DB 1001 GluLeuLeuTyrGlyGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1020  
QY 3061 GTGCTGCTTTCCAGCCCTTTGTTCTTAGGCTCAAGGCACCTTGAGTGGGTGAGACAACGA 3120  
DB 1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
QY 3121 GCGTGGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGAGACTTGCAGATCCCATCGT 3180  
DB 1041 AlaTrpLysGlnLeuArgGluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArg 1060  
QY 3181 TTCAAAGTGGAGATTTCAGTCTAGTTAGACGCCCGTCGAGGAAACCTTCGAGACTCGG 3240  
DB 1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
QY 3241 TGAAGGCGCTTATCTCGTACTTTTGCACACACACCGCTGGAAGTGAAGTGAAGGAATC 3300  
DB 1081 TrpLysGlyProTyrLeuValLeuLeuThrProThrAlaValLysValGluGlyIle 1100  
QY 3301 TCCACCTGGATCCATGTCATCCACGTTTAAACCCGCGCCACCTCCCGATTGGGGTGGAAA 3360  
DB 1101 SerThrTrpIleHisAlaSerHisValLysLeuAlaProProAspSerGlyTrpArg 1120  
QY 3361 GCGGAAAGACTGAAATATCCCTTAAGCTTCGCTTCCATCGGTGGTTCCTTACTCTGTC 3420  
DB 1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsn 1140  
QY 3421 AATAAC 3426  
DB 1141 AsnAsn 1142

RESULT 10  
Q90RL6\_9GAMR  
ID Q90RL6\_9GAMR PRELIMINARY; PRT; 1195 AA.  
AC Q90RL6\_9GAMR  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polymerase (Fragment).  
GN Name=pol;  
OS Porcine endogenous type C retrovirus.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Gamma-retrovirus.  
OX NCBI\_TaxID=70540;  
RN [1]  
RP NUCLEOTIDE SEQUENCE  
RX MEDLINE=21256017; PubMed=11356953;  
RX DOI=10.1128/JVI.75.12.5465-5472.2001;  
RA Krach U., Fischer N., Czaderna F., Toenjes R.R.;  
RT "Comparison of replication-competent molecular clones of porcine  
endogenous retrovirus class A and class B derived from pig and human  
cells";  
RL J. Virol. 75:5465-5472(2001).  
CC -/- SIMILARITY: Contains 1 reverse transcriptase domain.  
DR EMBL; AJ293657; CAC39626.1; -; Genomic DNA.  
DR HSSP; P03355; 116J.  
DR MSROPS; A02\_020; -;  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR001969; Pept\_Asp\_AS.  
DR InterPro; IPR009007; Pept\_AspArtc\_cat.  
DR InterPro; IPR001995; PeptIdase\_A2\_cat.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR012337; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS00994; INTEGRASE; 1.  
DR PROSITE; PS06879; RNASE\_H; 1.  
DR PROSITE; PS08878; RT\_POL; 1.  
KW Aspartyl protease; Hydrolase; Protease.  
FT NON\_TER 1  
SQ SEQUENCE 1195 AA; 133770 MW; 999F1524DE0BFFA4 CRC64;

## Alignment Scores:

Pred. No.:	0	Length:	1195
Score:	5861.00	Matches:	1108
Percent Similarity:	98.5%	Conservative:	17
Best Local Similarity:	97.0%	Mismatches:	17
Query Match:	94.8%	Indels:	0
DB:	2	Gaps:	0

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q90RL6\_9GAMR (1-1195)

QY	1	ATGGGTGCACAGGGCAACACAGATATCCATGACTACCCGAGAACAGTTGACTTTGGGA	60
Db	50	MetGlyAlaThrGlyGlnArgGlnTrpProTrpThrThrArgGthValAspLeuAla	69
QY	61	GTGGGACGGGTAAACCACCTCGTTTCTGCTATACCTGAGTGCACCGACCCCTCTTAGGT	120
Db	70	ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly	89
QY	121	AGAGACTTATTGACCAAGATGGGAGCACAAATTTCTTTTGACCAAGGAAACCAAGATG	180
Db	90	ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal	109
QY	181	TCTGCAAAATAACAACTTATCACTGTGTGACCTCCCAATTAGATGACGAATATCGACTA	240
Db	110	SerValAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyArgLeu	129
QY	241	TACTCTCCCTAGTAAGCCCTGATCAAAATATACAAATTTCTGGTTGGAAACAGTTTCCCAA	300
Db	130	TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGln	149
QY	301	GCCTGGGAGAAAACCGCAGGAGATGGGTTGGCAAGAGAGTTCCCCCAAGATTATTCAA	360
Db	150	AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln	169
QY	361	CTGAAGGCCAGTGCCACACCACTGTCAGTCAGACAGTACCCCTTGAGTAAAGAGCTCAA	420
Db	170	LeuLysAlaSerAlaThrProValSerValArgGlnTrpProLeuSerArgGluAlaArg	189
QY	421	GAAGGAATTCGGCCGTCATGTCCAAAGATTAATCCAACAGGGCATCTTAGTTCTCTGCCAA	480
Db	190	GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln	209
QY	481	TCTCCCTGGAATACTCCCTGCTACCGTTAGAAAGCCTGGGACTAATGACTATCGACCA	540
Db	210	SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyArgPro	229
QY	541	GTACAGGACTTGAAGAGGTCAATAAACGGGTGCAGGATATACACCAACAGTCCCGAAC	600

Db	230	ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn	249
QY	601	CTTTATAACCTCTTGTGTGTCTCCACCCCAACGGAGCTGGTATACAGTATTGGACTTA	660
Db	250	ProTyAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyThrValLeuAspLeu	269
QY	661	AAGGATGCCTTCTCTCTGAGATTACACCCCACTAGCCAAACCACTTTTTCCTTCGAA	720
Db	270	LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu	289
QY	721	TGGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTCCCAAGGG	780
Db	290	TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly	309
QY	781	TTCAAGAACTCCCGACATCTTTTGAGAAGCCCTACACAGAGACCTGGCCCACTTCAGG	840
Db	310	PheLysAsnSerProThrPhePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg	329
QY	841	ATCCAACACCTCAGGTGACCCCTCCTCAGTACCTGGATGACCTGCTTCTGGCGGAGCC	900
Db	330	IleGlnHisProGlnValThrLeuLeuGlnTrpValAspAspLeuLeuAlaGlyAla	349
QY	901	ACCAACACGAGTCTCTTAGAAGGCACGAAGGCACCTACTGCTGGAATTGTCTGACTAGGC	960
Db	350	ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly	369
QY	961	TACAGAGCTCTGCTAAGAGGCCAGATTTTGAGGAGAGAGGTAAACATATCTGGGGTAC	1020
Db	370	TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyLeuGlyTy	389
QY	1021	AGTTTGGGGACGGGACGAGTGTGACGAGGACGCGAAGAAACCTGTAGTCCAGATA	1080
Db	390	SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle	409
QY	1081	CCGGCCCCAACCAACAGCCAAACAAATGAGAGAGTTTTGGGGACAGCTGGGATTTGCAGA	1140
Db	410	ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg	429
QY	1141	CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTAACCAAGAAAAA	1200
Db	430	LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyProLeuThrLysGluLys	449
QY	1201	GGGGAATTTCTCTGGCTCCAGCACACAGAACGATTTGTATGCTATCAAAAAGGCCCTG	1260
Db	450	GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu	469
QY	1261	CTGAGCGCACCTGCTCTGGCCCTCCCTGACATAACTAAACCTTTTACCTTTATGTGGAT	1320
Db	470	LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTy-ValAsp	489
QY	1321	GAGGTAAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTTAGACCCATGAGAGAGACT	1380
Db	490	GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro	509
QY	1381	GTCCCTCACCTGTCAAAGAAGCTCGATCCTGTAGCTAGTGTGGCCCATATGCTGAAG	1440
Db	510	ValAlaTyLeuSerLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLys	529
QY	1441	GCTATCGCAGCTGTGGCCCATACTGCTCAAGACGCTGACAAATTTGACTTTGGGACAGAT	1500
Db	530	AlaIleAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn	549
QY	1501	ATAACTGTAATAGCCCCCATGCAATTGGAGAACATCGTTTCGGCAGCCCCCAGACCGATGG	1560
Db	550	IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp	569
QY	1561	ATGACCAACGCCCATGACCCACTATCAAAAGCTGTCTTCTCACAGAGAGGTACACGTTT	1620
Db	570	MetThrAsnAlaArgMetThrHisTyGlnSerLeuLeuLeuThrGluArgValThrPhe	589
QY	1621	GCTCCACGACCGCTCTCAACCTGCCCACTCTTCTGCTGAAGAGACTGTATGAACAGTG	1680
Db	590	AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal	609

QY 1681 ACTCATGATGCCATCAACTATTGATTGAGGAGACTGGGTCCGCAAGACCTTTACAGAC 1740  
DB ThrHisaspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 629  
QY 1741 ATACCGCTGACTGGAGAGTCTTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGA 1800  
DB IleProLeuThrGlyValLeuLeuThrTrpPheThrAspGlySerSerTyrValValGlu 649  
QY 1801 GGTAAAGAGATGGCTGGGGCGGCTGGTGGACGGGACCCACGATCTGGGGCCAGCAGC 1860  
DB GlyLysLysMetAlaGlyAlaValValaspGlyThrArgThrIleTrpAlaSerSer 669  
QY 1861 CTGCGCGAAGAACTTCAGCAACAAAGGCTGAGTCACTGGCCCTCAGCAAGCTTTGCGG 1920  
DB LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 689  
QY 1921 CTGCGCGAAGGAAATCATATAAATTTATACGACAGCAGGATGCTTTGCCACTGCA 1980  
DB LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 709  
QY 1981 CACGTACATGGGGCATCTATAAACAAGGGGTTGCTTACTCAGCAGGAGGAGAAATA 2040  
DB HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 729  
QY 2041 AAGAACAAAGAGGAAATCTTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCT 2100  
DB LysAsnLysGluGluLeuLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 749  
QY 2101 ATTATACATCTCTCGACATCAGAAAGCTTAAGATCTCATATCCAGAGAAACCATG 2160  
DB IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 769  
QY 2161 GCTCACCGGTTGCAAGCAGCAGCCAGGCTGTAAACCTTCTGCTTAAATAGAATG 2220  
DB AlaaspArgValAlaLysGlnAlaGlnAlaValAsnLeuLeuProIleIleGluThr 789  
QY 2221 CCCAAAGCCCAAGACCCAGACGACGACAGTACCCCTAGAGACTGGCAGAGATAAAAAAG 2280  
DB ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 809  
QY 2281 ATAGACCATCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATC 2340  
DB IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle 829  
QY 2341 CTGCCCCACAAAGAGGTTAGATATGTCCAAGATACATGCTTAACCCACTAGGA 2400  
DB LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 849  
QY 2401 ACTAAACACCTGCAGCAGTGGTGCAGAACATCCCTTATCATGTTCTGAGGCTACCAAGGA 2460  
DB ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuValArgLeuProGly 869  
QY 2461 GTGGCTGACTCGGTGGTCAACATGTTGTGGCTCCGACCTGCTGTTAATGTAATCTCTCC 2520  
DB ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 889  
QY 2521 AGAATGCTCCAGGGAAGACTAAGGGGAAGCCACCGCGCTCACTGGGAGTGCAC 2580  
DB ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 909  
QY 2581 TTCCTCAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGTTTGTGTAGACACC 2640  
DB PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 929  
QY 2641 TTTTCAGGTGGGTAGAGCTTATCTCTAAGAAAGAGACTTCAACCGGTGGTGGCTAAA 2700  
DB PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 949  
QY 2701 AAAATACCTGAAGAAATTTTCCAAAGATTTGGATACCTAAGGTAAATAGGCTCAGCAAT 2760  
DB LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 969

QY 2761 GGTCCAGCTTTTGTGGCCAGGTAAGTCAGGACTGGCCAGATATATGGGGATTGATGG 2820  
DB GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyLeuAspTrp 989  
QY 2821 AAATGTCATTTGTGCATACAGACCCCAAGCTCAGACAGGTAGAGAGATGAATAGAACC 2880  
DB LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 1009  
QY 2881 ATTAAGAGACCTTTACTTAAATTCACCGGAGACTGGCGTAAATGATTGATAGTACTCTC 2940  
DB IleLysGluThrLeuThrLysLeuThrGluThrGlyIleAsnAspTrpIleAlaLeu 1029  
QY 2941 CTGCCCCCTTTGTCTTTTGTAGGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCTAT 3000  
DB LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1049  
QY 3001 GAATTAATCTACCGGGGACCCCGCCCATTTGGTAGAAAATGCTTCTGTACATAGTCTGCAC 3060  
DB GluLeuLeuTyrGlyGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1069  
QY 3061 GTGCTGCTTTCCAGACCTTTGTTCTTAGGCTCAAGGCACCTTGAGTGGGTGAGACAACGA 3120  
DB ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1089  
QY 3121 GCGTGGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGGAGACTTGCAGATCCCACTCGT 3180  
DB AlaTrpLysGlnLeuArgGluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArg 1109  
QY 3181 TTCCAAGTGGAGATTCACTGCTAGCTTAGAGCCACCGCTGAGGAAACCTCGAGACTCGG 3240  
DB PheGlnValGlyAspSerValTyrValArgArgHisAlaGlyAsnLeuGluThrArg 1129  
QY 3241 TGAAGGGCCCTTATCTGCTACTTTTGACCACCAACCGCTGTGAAGTCTGAAGGAATC 3300  
DB TrpLysGlyProTyrLeuValLeuLeuThrProThrAlaValLysValGluGlyIle 1149  
QY 3301 TCCACCTGGATCCATGCATCCACGTTAAACCGCGCCACCTCCCGATTTCGGGTGGAAA 3360  
DB SerThrTrpIleHisAlaSerHisValLysLeuAlaProProAspSerGlyTrpArg 1169  
QY 3361 GCGAAAGACTGAATAATCCCTTAAGCTTCGCTCCATCGGTGGTTCCTTACTCTGTC 3420  
DB AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsn 1189  
QY 3421 AATAAC 3426  
DB 1190 AsnAsn 1191  
RESULT 11  
Q9Q1X3\_9GAMR  
ID Q9Q1X3\_9GAMR PRELIMINARY; PRT; 2376 AA.  
AC Q9Q1X3;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Type C proviral gag, pol and env genes and LTR (class B, clone 43).  
OS Porcine endogenous retrovirus.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC GammaRetrovirus; 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=Type C;  
RX MEDLINE=20219394; PubMed=10756014;  
DOI=10.1126/JVI.74.9.4028-4038.2000;  
RA Czaderna F., Fischer N., Boller K., Kurth R., Toenjes R.R.;  
RT "Establishment and characterization of molecular clones of porcine  
endogenous retroviruses replicating on human cells";  
RL J. Virol. 74:4028-4038(2000).  
DR ENBL; AJ133818; CAB65341.1; -; Genomic\_DNA.  
DR HSSP; P03385; IMOF.  
DR SMR; Q9Q1X3; 7-98, 2225-2277.

DR MEROPS; A02.020; --  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006310; F:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0019068; P:viral assembly; IEA.  
DR InterPro; IPR002050; Env polyprotein.  
DR InterPro; IPR008981; FmuYrecept-bind.  
DR InterPro; IPR003036; Gag p30.  
DR InterPro; IPR000840; G retro matrix.  
DR InterPro; IPR001995; Peptidase A2 cat.  
DR InterPro; IPR001969; Pept Asp\_AS\_  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVtse.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF01140; Gag MA; 1.  
DR Pfam; PF02093; Gag p30; 1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR Pfam; PF00429; TLV coat; 1.  
DR Pfam; PF00098; zf\_CCHC; 1.  
DR SMART; SM00343; Znf\_C2HC; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS00879; RNASE H; 1.  
DR PROSITE; PS00158; ZF\_CCHC; 1.  
FT CHAIN 1 524 Gag protein.  
FT CHAIN 525 1761 Pol protein.  
FT CHAIN 1720 2376 Env protein.  
SQ SEQUENCE 2376 AA; 266550 MW; C6D16E368BA49ABP CRC64;

Alignment Scores:  
Pred. No.: 0 Length: 2376  
Score: 5861.00 Matches: 1108  
Percent Similarity: 98.5% Conservative: 17  
Best Local Similarity: 97.0% Mismatches: 17  
Query Match: 94.8% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q9Q1X3\_9GAMR (1-2376)

Qy 1 ATGGGTGCCACAGGGCAACACAGTATCCATGGACTACCCGGAACAGTGTGACTTGGGA 60  
Db 574 MetGlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAspLeuAla 593

Qy 61 GTGGGAGGGTAACCCACTCGTTCTGGTCATACCTGAGTGCCCGACGCCCTCTTAGGT 120  
Db 594 ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly 613

Qy 121 AGAGACTATTATGACCAAGATGGGAGCAAAATTTCTTTGAAACAAGGMAACCAAGAGTG 180  
Db 614 ArgaspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal 633

Qy 181 TCTCAATAACAAACCTTATCACTGTGTGACCTTCCAAATTTAGATGACGAATATCGACTA 240  
Db 634 SerValAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu 653

Qy 241 TACTCTCCCTAGTAAGCCTGATCAAAATATCAATTCCTGTTGGACAGTTTCCCAAA 300  
Db 654 TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGluGlnPheProGln 673

Qy 301 CCCTGGGCAGAAACCGCAGGAGTGGTTTGGCAAGCAAGTTCCTCCCAACAGATTATTCAA 360  
Db 674 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProGlnValIleGln 693

Qy 361 CTGAAGGCCAGTGCCACACAGTGTCTCAGTCAGAGTACCCCTTGAGTAAAGAAGCTCAA 420  
Db 694 LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaArg 713

Qy 421 GAAGGAATTCGGCGCATGTCCAAAGATTAAATCAACAGAGGCATCCTAGTTCCTGTCCAA 480  
Db 714 GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 733

Qy 481 TCTCCCTGGAAATCTCCCTGCTTACCGGTTAGAAAGCCTGGGACTAATAGACTATCGACCA 540  
Db 734 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 753

Qy 541 GTACGAGACTTGAGAGAGGTCAATACAGGCTGAGGATATACACCAACAGTCCCGAAC 600  
Db 754 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 773

Qy 601 CTTTATAACCTCTTTGTGTCTCTCCACCCCAACGAGCTGGTATACAGTATTGGACTTA 660  
Db 774 ProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeu 793

Qy 661 AAGGATCCCTTCTCTCCCTGAGATTACACCCACTAGCCAAACCACTTTTTCCTTCGAA 720  
Db 794 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 813

Qy 721 TGGAGAGATCCAGTACGAGGAGNACCGGCGAGCTCACCTGGACCCGACTGCCCAAGGG 780  
Db 814 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 833

Qy 781 TTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCCAACTTCAGG 840  
Db 834 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 853

Qy 841 ATCCAACACCCCTCAGGTGACCCCTCTCCAGTACGTGAGTACCTGCTGGAAATGCTGACCTAGGC 900  
Db 854 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAla 873

Qy 901 ACCAAACAGGACTGCTTAGAAGGCACCAAGCCACTACTGCTGGAAATGCTGACCTAGGC 960  
Db 874 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 893

Qy 961 TACAGAGCCTCTGTGAAGAGGCCACAGATTTCAGGAGAGAGGTAACATCTTGGGGTAC 1020  
Db 894 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 913

Qy 1021 AGTTTGGCGACGGGACGAGATGCTGACGGAGCAGGAGCAAGAACTGTAGTCCAGATA 1080  
Db 914 SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 933

Qy 1081 CCGGCCCCAACCCACAGCCCAACAAATCAGAGAGTTCCTTGGGACAGCTGGATTTTGCAGA 1140  
Db 934 ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 953

Qy 1141 CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTAACCAAGAAAAA 1200  
Db 954 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLys 973

Qy 1201 GGGGAATTCTCTGGGCTCTGAGCACACAGAAGCATTTGTATGCTATCAAAAAGCCCTG 1260  
Db 974 GlyGluPheSerTrpAlaProGluHisGlnLysThrPheAspAlaIleLysLysAlaLeu 993

Qy 1261 CTGAGCCGACCTGCTCTGGCCCTCCCTGACGTACTATAACCTTTTACCTTTATGTGAT 1320  
Db 994 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 1013

Qy 1321 GAGCGTAAGGAGTAGCCCGGGAGTTCCTTAACCCCAACCCCTTAGGACCATGAGAGAGCCT 1380  
Db 1014 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 1033

Qy 1381 GTCCCTTACCTGTCAAAGAAGCTCGATCTGTAGCCAGTGGTTGGCCCATATGCTGAAG 1440  
Db 1034 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLys 1053



1441 GCTATCGCAGCTGTGGCCATCACTACTGCTCAAGGACCGCTGACAAATTTGACTTTGGGACAGAAAT 1500  
1054 AlalileAlaAlaValAlaAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 1073  
1501 ATAACCTGTAATAGCCCGCCCATCTGCTATTGGAGAACATCGTTTCGGCAGCCCGCCAGACCGATGG 1560  
1074 IleThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnProProAspArgTrp 1093  
1561 ATGACCAACCGCCCGCATGACCCCATCTATCAAGCCCTGCTTCTCACAGAGAGGGTCAACGCTTC 1620  
1094 MetThrAsnAlaArgMetThrHisGlnSerLeuLeuLeuThrGluArgValThrPhe 1113  
1621 GCTCCACGAGCGCTCTCAACCCCTGCTCACTCTTCTGCTGAGAGACTGATGAACAGATG 1680  
1114 AlaProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 1133  
1681 ACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGTCGCAAGGACCTTACAGAC 1740  
1134 ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 1153  
1741 ATACCGCTGACTGGAGAAGTCTAACTTGTCTCACTGACGGAAGCAGCTATGTGGTGAA 1800  
1154 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 1173  
1801 GGTAAAGAGGATCGCTGGGGCGCGGTGTGTGACCGGACCGCCAGCATCTGGGCCAGCAGC 1860  
1174 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrHisThrIleTrpAlaSerSer 1193  
1861 CTGCGCGAAGAACTTCAGACACAAAGCTGAGCTCATGCGCTCAGCAGCAGCTTTCGGG 1920  
1194 LeuProGluGlyThrSerAlaGlnLysAlaGlnLeuMetAlaLeuThrGlnAlaLeuArg 1213  
1921 CTGCGCGAAGGAAATCATATAACATTTATACGACAGCAGGATGCTTTCGACCTGCA 1980  
1214 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 1233  
1981 CAGTCATACGGGCGCATCTATAACAAAGGGGGTGTCTTACCTCAGCAGGAGGAGAAATA 2040  
1234 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 1253  
2041 AAGAACAAAGAGAAATTTCTAAGCTATTAGAGCCGTACATTTACCAGAAAGCGCTAGCT 2100  
1254 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 1273  
2101 ATTATACACTGTCTGTGACATCAAGAACTAAGATCTCATATCCAGAGGAAACCCAGATG 2160  
1274 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 1293  
2161 GCTGACCGGTTGCCAAGCAGCGCGCCCGGGTGTAACTTCTGCTTAAATAGAAATG 2220  
1294 AlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeuLeuProIleIleGluThr 1313  
2221 CCCAAAGCCCGAAGACCCAGACGACGATACACCTTAGAAGCTGGCAAGAGATATAAAAG 2280  
1314 ProLysAlaProGluProArgArgLysGlnLysLeuGluAspTrpGlnGluIleLysLys 1333  
2281 ATAGACCGATTCTGTGAGACTCCGGAAGGACCTGTATACCTCAGATGCGGAAGGAATC 2340  
1334 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle 1353  
2341 CTGCCCCCAAGAGAGGGTTAGAAATATGTCCACAGATACATCGTCTAACCCACCTAGGA 2400  
1354 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 1373  
2401 ACTAAACACCTGCAGCAGTTGCTCAGNACATCCCTTATCATCTTCTGAGGCTTACCAGGA 2460  
1374 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 1393  
2461 GTGGCTGACTCGGTGGTCAACAACTTGTGTCCTGCGCAGCTGGTGAATGCTTAATGCTTCTTC 2520  
1394 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 1413  
2521 AGAATGCTCCAGGGAAGAGACTTAAGGGGAAGCCACCCAGCGCTCACTGGGAAGTGGAC 2580

1414 ArgValProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 1433  
2581 TTCACTCAGGTAAAGCCGGCTAAATATACGGAACAAATACCTATTGTTGTTTGTAGACACC 2640  
1434 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 1453  
2641 TTTTCAGATGGGTAGAGGCTTATCTTACTAAAGAAAGAGACTTCAACCGCTGGTGGCTAAA 2700  
1454 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 1473  
2701 AAAATACCTGAAGAAATTTTCCAGATTTCGAATATACCTAAGGTAAATAGGTCAGACAAAT 2760  
1474 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 1493  
2761 GGTCCACGCTTTGTTCGCCAGGTAAAGTCAGGACTGCGCAAGATATTGGGGATTGATTGG 2820  
1494 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 1513  
2821 AAATGCTATTGTGCATACAGACCCCAAGCTCAGGACAGGCTAGAGAGATGAATAGAAC 2880  
1514 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnLysThr 1533  
2881 ATTAAGAGACCTTACTTAATTTGACCCGCGAGAGACTGCGCTTAATGATTGATAGCTCTC 2940  
1534 IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeu 1553  
2941 CTCGCCCTTGTGCTTTTAGGTTAGGAACACCCCTGGACAGCTTGGGCTGACCCCTCAT 3000  
1554 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1573  
3001 GAATTAATCTACCGGGGACCCCGCTTCTGTTAGAAATTTGCTCTGACATAGTCTGCTGAC 3060  
1574 GluLeuLeuTyrGlyGlyProProProLeuValGluIleAlaSerValHisSerAlaAsp 1593  
3061 GTGCTGCTTTCAGCGCTTTGTTCTCTAGGCTCAAGGCACTTGAAGTGGTGAGACAACGA 3120  
1594 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1613  
3121 GCGTGGAGGCAACTCCGGGAGGCTACTCAGGAGGAGGACTTGCAGATCCACATCGT 3180  
1614 AlaTrpLysGlnLeuArgGluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArg 1633  
3181 TTCACAAAGGAGATTCAGTCTACGTTAGACGCGCCGCTCAGGAAACCTCGAGACTCGG 3240  
1634 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1653  
3241 TGAAGGGCCCTTATCTCGTACTTTTGAACCAACCGCTGTGAAGTCTGAAGGAATC 3300  
1654 TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1673  
3301 TCCACCTGGATCCATCCACCTCCACGTTAAACCGCGCCACTCCGATTCGGGGTGGAAA 3360  
1674 SerThrTrpIleHisAlaSerHisValLysLeuAlaProProProAspSerGlyTrpArg 1693  
3361 GCCGAAAGACTGAAATCCCTTAAAGCTTCGCTCCATCGCTGGTTCCTTACTCTGTC 3420  
1694 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsn 1713  
3421 ATAAC 3426  
1714 AsnAsn 1715  
RESULT 12  
Q8UM99\_9GAMR PRELIMINARY; PRT; 1144 AA.  
ID Q8UM99\_9GAMR  
AC Q8UM99;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Poi protein.  
GN Name=pol;  
OS Porcine endogenous retrovirus.

OC Viruses; Retrovirdae; Retroviridae; Gammatetrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN (1)  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=20219394; PubMed=10756014;  
RX DOI=10.1128/JVI.74.9.4028-4038.2000;  
RX Czauderna F., Fischer N., Boller K., Kurth R., Toenjes R.R.;  
RA "Establishment and characterization of molecular clones of porcine  
RT endogenous retroviruses replicating on human cells.";  
RL J. Virol. 74:4028-4038(2000).  
RN (2)  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=21851083; PubMed=11861838;  
RX DOI=10.1128/JVI.76.6.2714-2720.2002;  
RX Niebert M., Rogel-Gaillard C., Chardon P., Toenjes R.R.;  
RA "Characterization of Chromosomally Assigned Replication-Competent Gamma  
RT Porcine Endogenous Retroviruses Derived from a Large White Pig and  
RT Expression in Human Cells";  
RL J. Virol. 76:2714-2720(2002).  
DR EMBL; AJ279056; CAC82502.1; -, Genomic\_DNA.  
DR HSSP; P03355; 116J.  
DR MEROPS; A02.020; -.  
DR GO; GO:0003677; P:DNA binding; IEA.  
DR GO; GO:0004523; P:ribonuclease H activity; IEA.  
DR GO; GO:0003723; P:RNA binding; IEA.  
DR GO; GO:0003964; P:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR PROSITE; PS00879; RNASE H; 1.  
SQ SEQUENCE 1144 AA; 127926 MW; 5F3AA5AEFB98038 CRC64;  
  
Alignment Scores:  
Pred. No.: 0 Length: 1144  
Score: 5840.50 Matches: 1107  
Percent Similarity: 97.8% Conservative: 12  
Best Local Similarity: 96.8% Mismatches: 24  
Query Match: 94.5% Indels: 1  
DB: 2 Gaps: 1  
  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q8UM99\_9GAMR (1-1144)  
  
QY 1 ATGGGTGCCACAGGCGCAACAGTATCCATGGACTACCCGAGAACAGTTGACTTGGGA 60  
DB 1 MetGlyAlaThrGlyGlnGlnGlnTyrProThrThrArgA-gThrValAspLeuGly 20  
  
QY 61 GTGGGACGGGTAAACCACTCGTTCTGGTCATACCTGAGTGCACCGACCCCTCTTAGGT 120  
DB 21 ValGlyArgValThrGlnSerPheLeuGlyIleProGluGlyProGluProLeuLeuGly 40  
  
QY 121 AGAGACTTATTGACCAAGATGGGAGCAAAATTTCTTTTGAACAGGGAACCGAAGTG 180  
DB 41 ArgAspLeuLeuThrThrMetGlyArgGlnIlePhePheGluGlnGlyThrProGluVal 60  
  
QY 181 TCTGCAATAACAACCTATCACTGTGTGGACCTCCCAATTAGATGACGAATATCGACTA 240  
DB 61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu 80  
  
QY 241 TACTCTCCCTAGTAAAGCCTGATCAAAATATACAAATTTCTGGTTGGAAACAGTTTCCCAA 300  
DB 81 TyrSerSerLeuValLysProAspGlnAsnIleGlnPheThrLeuGluGlnPheProGln 100  
  
QY 301 GCCTGGGAGAAAACCGAGGAGATGGGTTGGCAAGCAAGTTCCCCCAAGTATTTCAA 360  
DB 101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 120  
  
QY 361 CTGAAGGCCAGTGCCACACCGAGTGTCACTGACAGACAGTACCCCTTGAGTAAAGAGCTCAA 420

Db 121 LeuLysAlaSerAlaAlaProValSerValArgGlnTyrProLeuSerLysGluAlaArg 140  
QY 421 GAAGGAATTCGGCGCATGTCCTCCAAAGATTAAATCCACAGGCGCATCTAGTTCCTGCCAA 480  
Db 141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 160  
QY 481 TCTCCCTGGGAATATCTCCCTGCTACCGGTTAGAAAGCCTGGGACTAATGACTATCGACCA 540  
Db 161 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 180  
QY 541 GTACAGCACTTGACAGAGGTCAATAAATACGGGTGCAGGATATACACCAACAGTCCCGAAC 600  
Db 181 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200  
QY 601 CCTTATTAACCTCTTGTGTGCTCTCCACCCCAACGAGCTGGGTATACAGTATTGGACTTA 660  
Db 201 ProTyrAsnLeuLeuCysAlaLeuProGlnArgSerTrpTyrThrValLeuAspLeu 220  
QY 661 AAGGATCGCTTCTCTGCTGAGATTACACCCACTAGCAACCACTTTTTCCTTCGAA 720  
Db 221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240  
QY 721 TGGAGAGATCCAGGTACGGGAAGAACCGGCGAGCTCACCTGGACCCGACTGCCCAAGGG 780  
Db 241 TrpArgAspProGlyAlaGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260  
QY 781 TTCAAGAACTCCCGACCATCTTTGAGAGACCCCTACACAGAGACCTGGCCAACTTCAGG 840  
Db 261 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280  
QY 841 ATCCAACACCTCAGGTGACCTCTCCTCAGTACGTGATGACCTGCTCTCGCGGGAGCC 900  
Db 281 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla 300  
QY 901 ACCAAACAGGACTCTTTAGAAAGGCACGAAGGCACACTACTGTGGAAATTTCTGACTAGGC 960  
Db 301 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 320  
QY 961 TACAGACCTCTGTAGAAAGGCCAGATTTCGAGGAGAGAGGTAAACATCTTGGGTAC 1020  
Db 321 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 340  
QY 1021 AGTTTGGGACGGGACGATGCTGACGAGGACCGAAGAAACCTAGTGTCCAGATA 1080  
Db 341 SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysArgThrValValGlnIle 360  
QY 1081 CCGGCCCAACCAACAGCCAAACAAATGAGAGAGTTCCTTGGGACAGCTGGGATTTTGCAGA 1140  
Db 361 ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 380  
QY 1141 CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCTTACCCGCTAACCAAGAAAAA 1200  
Db 381 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLys 400  
QY 1201 GGGGAATTCCTGGGCTCCTGAGCACCAGAGGATTTGATGCTATCAAAAAGCCCTG 1260  
Db 401 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420  
QY 1261 CTGAGCGCACCTGCTGGCCCTCCTGACGTAATAACCTTTTACCCCTTATGTGGAT 1320  
Db 421 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 440  
QY 1321 GAGGTAAAGGAGTAGCCCGGGAGTTTAAACCAACCCCTAGACCCATGAGGAGACCT 1380  
Db 441 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 460  
QY 1381 GTCCCTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCTGAAG 1440  
Db 461 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTyrProValCysLeuLys 480  
QY 1441 GCTATCGCAGCTGTGGCCATATCTGGTCAAGGACCTGACAAATGACTTTTGGACAGAT 1500

Db 481 AlalleAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 500  
Qy 1501 ATAAGTGAATAGAGCCCATGCAATGAGAACATCGTTCCGACAGCCCCACAGACGATGG 1560  
Db 501 IleThrValIleAlaProHisAlaLeuGlnAsnIleValArgGlnProProAspArgTrp 520  
Qy 1561 ATGACCAAGCCCGGATGACCACTATCAAGACCTGCTTCTCAGAGAGGGTCAAGCTTC 1620  
Db 521 MetThrAsnAlaArgMetThrHisTyrglnSerLeuLeuThrGluArgValThrPhe 540  
Qy 1621 GCTCCACGAGCGCTCTCAACCCCTGCCACTCTCTGCTGAGAGAGACTGTAACCCAGTG 1680  
Db 541 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 560  
Qy 1681 ACTCATGATGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACTTACAGAC 1740  
Db 561 ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 580  
Qy 1741 ATACCGCTGACTGGAGAAGTCTAACCTGGTTCTACTGACGGAAGCAGCTATGTGGTGA 1800  
Db 581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrrValValLys 600  
Qy 1801 GGTAAAGAGGATGGCTGGGGCGGCTGTGTGGACGGGACCCGACGATCTGGCCGAGCAG 1860  
Db 601 GlyLysArgMetAlaGlyProProValValAspGlyThrArgThrIleTrpAlaSerSer 620  
Qy 1861 CTGCGCGAAGGAATCTCAGCAACAAAGGCTCAGCTCATGGCCCTCAGCAAGCTTTGCGG 1920  
Db 621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
Qy 1921 CTGCGCGAAGGAATCATTAACATTATACGACAGCAGGTATGCTTTGCGACTGCA 1980  
Db 641 LeuAlaGluGlyLysSerIleAsnIleTyrrThrAspSerArgTyrrAlaPheAlaThrAla 660  
Qy 1981 CACGTACATGGGCATCTATAACAAGGGGTGCTTACCTCAGCAGGGAGGAGAAATA 2040  
Db 661 HisValHisGlyAlaIleTyrrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
Qy 2041 AAGAAACAAAGAGAAATCTTAAGCTATTAGAAGCGCTACATTACAAAGAGCTAGCT 2100  
Db 681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 700  
Qy 2101 ATTATACATCTCTCGACATCAGAAAGCTAAAGATCTCATATCCAGAGAAACCATG 2160  
Db 701 IleIleHisCysProGlyHisGlnLysAlaLysAspProIleSerArgGlyAsnGlnMet 720  
Qy 2161 GCTGACCGGCTGCGACGAGGACCCGAGGTGTTAACTTCTGCTTATATAAGAAATG 2220  
Db 721 AlaAspArgValAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProMetIleGluThr 740  
Qy 2221 CCCAAAGCCCGAAGCCAGACGACGTACACCCCTAGAAGACTGGCAAGAGATAAAAAAG 2280  
Db 741 ProLysAlaProGluProGlyArgGlnTyrrThrLeuGluAspTrpGlnGluIleLysLys 760  
Qy 2281 ATAGACAGATTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGGAATC 2340  
Db 761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrrThrSerAspGlyLysGluIle 780  
Qy 2341 CTGCCCCACAAAGAGGGTTAGATATGTCCACAGATACATGCTTAACCCACCTAGGA 2400  
Db 781 LeuProHisLysGluGlyLeuGluTyrrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
Qy 2401 ACTAAACACTGACGAGTGGTGCAGAACATCCCTTATCATGTCTGAGCGCTACACAGGA 2460  
Db 801 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrrHisValLeuArgLeuProGly 820  
Qy 2461 GTGCGTCACTCGGTGGTCAACATATGTGTGCGCTCCGACGCTGGTGTATGTCTTCTTCC 2520  
Db 821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
Qy 2521 AGAATGCTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGCGCTCACTGGGAAGTGGAC 2580  
Db 841 ArgIleProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860

Qy 2581 TTCACTGAGGTAAAGCCGGCTAAATACGAAACAAATACCTATTGGTTTGTAGACACC 2640  
Db 861 PheThrGluValLysProAlaLysTyrrGlyAsnLysTyrrLeuLeuValPheValAspThr 880  
Qy 2641 TTTTCAGAGTGGTAGAGCTTATCTTACTTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 2700  
Db 881 PheSerGlyTrpValGluAlaTyrrProThrLysLysGluThrSerThrValValAlaLys 900  
Qy 2701 AAAATACCTGGAAGAAATTTTCCAAAGATTGGAATACCTAAGGTAAATAGGCTCAGACAA 2760  
Db 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
Qy 2761 GGTCCAGCTTTTGTGTCGCCAGGTAAAGTACAGGACTGGCCCAAGATATTGGGATTCATTGG 2820  
Db 921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
Qy 2821 AAATGCTGATTGTGCATACAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAAC 2880  
Db 941 LysLeuHisCysAlaTyrrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
Qy 2881 ATTAAGAGAGACCTTACTATAATTGACCGCGAGACTGGCGTTAATGATTGGATAGCTCTC 2940  
Db 961 IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeu 980  
Qy 2941 CTGCGCTTTGCTGCTTTTAGGGTTAGGAACAACCCCTCGACAGTTTGGCTCAGCCCTTAT 3000  
Db 981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrr 1000  
Qy 3001 GAATTACTCTACGGGGGACCCCGCCCATTTGGTAGAAATTCCTCTGTACATAGTCTGCAC 3060  
Db 1001 GluLeuLeuTyrrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisSerAlaAsp 1020  
Qy 3061 GTGCTGCTTTCCAGACCTTTGTTCTTAGCTCAAGGCACCTTGAGTGGGTGAGCAACGA 3120  
Db 1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
Qy 3121 GCGTGGAGGCAACTCCGGGAGGCGCTACTCAGGAGGAGAGACTTGCAGATCCACATCGT 3180  
Db 1041 AlaTrpLysGlnLeuArgGluAlaTyrrSer---GlyGlyAspLeuGlnValProHisArg 1059  
Qy 3181 TTCAGTGGAGATTCAGTCTCAGTTAGACGCCCGCTCAGGAAACCTCGAGACTCGG 3240  
Db 1060 PheGlnValGlyAspSerValTyrrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1079  
Qy 3241 TGGNAGGGCCCTTATCTCGTACTTTTCACACACCAACGCGCTGGAAGTCGAAGGAATC 3300  
Db 1080 TrpLysGlyProTyrrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1099  
Qy 3301 TCCACCTGGATCCATGCATCCACGTTTAAACCCGCGCCACTCCCGATTCGGGGTGGAAA 3360  
Db 1100 ProThrTrpIleHisAlaPheHisValLysProAlaProProSerAspSerGlyTrpLys 1119  
Qy 3361 GCCCAAGAGACTGAAATCCCTTAAAGCTTCGCTCCATCGCTGGTTCCTTACTCTGTC 3420  
Db 1120 AlaGlnLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrrSerVal 1139  
Qy 3421 AATAACCTCTCA 3432  
Db 1140 AsnAsnSerSer 1143

## RESULT 13

Q8UM96\_9GAMR  
ID Q8UM96\_9GAMR PRELIMINARY; PRT; 1147 AA.  
AC Q8UM96;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Pol protein.  
GN Name-pol;  
OS Porcine endogenous retrovirus.  
OC Viruses; Retrovirdae; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.

OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20219394; PubMed=10756014;  
RX DOI=10.1128/JVI.74.9.4028-4038.2000;  
RA Czauderna F., Fischer N., Boller K., Kurth R., Toenjes R.R.;  
RA "Establishment and characterization of molecular clones of porcine  
RT endogenous retroviruses replicating on human cells";  
RL J. Virol. 74:4028-4038(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21851083; PubMed=11861838;  
RX DOI=10.1128/JVI.76.6.2714-2720.2002;  
RA Niebert M., Rogel-Galliard C., Chardon P., Toenjes R.R.;  
RT "Characterization of Chromosomally Assigned Replication-Competent Gamma  
RT Porcine Endogenous Retroviruses Derived from a Large White Pig and  
RT Expression in Human Cells";  
RL J. Virol. 76:2714-2720(2002).  
DR EMBL: AJ279057; CAC82505.2; -, Genomic\_DNA.  
DR MEROPS: A02.020; --  
DR GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO: GO:0003677; F:DNA binding; IEA.  
DR GO: GO:0004523; F:ribonuclease H activity; IEA.  
DR GO: GO:0003723; F:RNA binding; IEA.  
DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO: GO:0006310; P:DNA recombination; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO: GO:0006278; P:protein-dependent DNA replication; IEA.  
DR InterPro: IPR001995; Peptidase\_A2\_cat.  
DR InterPro: IPR002156; RNaseH.  
DR InterPro: IPR001584; Rve.  
DR InterPro: IPR008477; RVTse.  
DR Pfam: PF000075; RNaseH; 1.  
DR Pfam: PF00665; rve; 1.  
DR Pfam: PF00077; RVP; 1.  
DR Pfam: PF00078; RVT\_1; 1.  
DR PROSITE: PS0175; ISP\_PROT\_RETROV; 1.  
DR PROSITE: PS0879; RNASE\_H; 1.  
SQ SEQUENCE 1147 AA; 128404 MW; C6374B0CE460901D CRC64;  
  
Alignment Scores:  
Pred. No.: 0 Length: 1147  
Score: 5831.00 Matches: 1104  
Percent Similarity: 98.4% Conservative: 20  
Best Local Similarity: 96.7% Mismatches: 18  
Query Match: 94.3% Indels: 0  
DB: 2 Gaps: 0  
  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q8UM96\_9GAMR (1-1147)  
QY 1 ATGGGTGCCAGGGCAACACAGTATCCATGGACTACCCGGAACAGTTGACTTGGGA 60  
DB 1 MetGlyAlaThrGlyGlnArgGlnTyrProTrpThrArgArgThrValAspLeuVal 20  
QY 61 GTGGGAGGGTAACCCACTCGTTCTGTCTATACCTGAGTCCCGACGCCCTCTTAGGT 120  
DB 21 ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly 40  
QY 121 AGAGACTTATTGACCAAGATGGGACACAAATTTCTTTTGAACAGGGAACCAAGATG 180  
DB 41 ArgAspLeuLeuThrLysMetGlyAlaGlnLeuSerPheGluGlnGlyArgProGluVal 60  
QY 181 TCTGCAATAACAAACCTATCACTGTGTGTGACCTCCATTAATTAGATGAGGAATATCGACTA 240  
DB 61 SerValAspAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu 80  
QY 241 TACTCTCCCTAGTAAGCCCTGATCAAAATATCAATTTCTGTGTGGACAGTTTCCCCAA 300  
DB 81 TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGlnPheProGln 100  
QY 301 GCCTGGGAGAAACCGAGGGATGGTTTGGCAAGCAAGTATCCCCCAACAAGTTATTCAA 360  
DB 101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProGlnValIleGln 120

QY 361 CTGAAGCCAGTGCACACAGTGTCTAGTCAGACAGTACCCCTTGAGTAAAGAGCTCAA 420  
DB LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaArg 140  
QY 421 GAAGGAATTCGGCGCATGTCCAAAGATTAATCAACAGGGCATCTTAGTTCCTGTCCAA 480  
DB GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyLeuValProValGln 160  
QY 481 TCTCCCTGGAAATCTCCCTGCTACCGGTTAGAAGCCTGGGACTAATGACTATCGACCA 540  
DB SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 180  
QY 541 GTACAGGACTTGTAGAGAGGTCAATAAACGGGTGAGGATATACACCAACAGTCCCAAC 600  
DB ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200  
QY 601 CCTTATAACCTCTTGTGTGCTCTCCCAACCCGAGCTGGTATACAGTATTGGACTTA 660  
DB ProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeu 220  
QY 661 AAGATGCTCTTCTGCTGAGATTACACCCACTAGCCACCAACCACTTTTTCCTTCGAA 720  
DB LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240  
QY 721 TGGAGAGATCCAGTACGGGAAGAACCGGCGAGCTCACCTGGACCCGACTGCCCAAGGG 780  
DB TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260  
QY 781 TTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCCACTTCAGG 840  
DB PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280  
QY 841 ATCCAAACCCCTCAGGTGACCTCTCTCAGTACGTGATGACCTGCTTCTGGCGGAGCC 900  
DB IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAla 300  
QY 901 ACCAAACAGGACTCTGTAGAGGACCAAGGCACACTACTGCTGGAATTTGTCTGACTAGGC 960  
DB ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuSerAspLeuGly 320  
QY 961 TACAGAGCCTCTGTAGAAGGCCCCAGATTTCGAGGAGAGAGTAACATCTTGGGTGAC 1020  
DB TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 340  
QY 1021 AGTTTGGGACGGGAGCGATGCTGACGGAGCAGGAGCAAGGAAACTGTAGTCCAGATA 1080  
DB SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 360  
QY 1081 CCGGCCCAACACACAGCAACAAATGAGAGAGTTTTTTGGGGACAGCTGGATTTTGACA 1140  
DB ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 380  
QY 1141 CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCTCTACTACCCGCTAACCAAGAAAA 1200  
DB LeuTrpIleProGlyPheAlaThrLeuAlaProLeuTyrProLeuThrLysGluLys 400  
QY 1201 GGGGAATTTCTCGGGCTCTGAGCACCAGAGGCATTTGATGCTATCAAAAGGCCCTG 1260  
DB GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420  
QY 1261 CTGAGCCCATCTCTGGCCCTCCCTGACCTAAACCCCTTTTACCTTTTATGTGGAT 1320  
DB LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 440  
QY 1321 GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCAACCCCTTAGGACCATGAGAGACCT 1380  
DB GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 460  
QY 1381 GTCGCCCTACCTGTCAAAGAGCTCGATCCTGTACCGAGTGGTTGGCCCATATGCTGAAG 1440  
DB ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLys 480

QY 1441 CCTATCGAGCTGGCGCATCTACTGGTCAAGGACGCTGACAAATTGACTTTGGGACAGAA 1500  
Db AlaileAlaValAlaileLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAen 500  
QY 1501 ATAAGTGAATAGCCCGCCATCGTATGGAGAACATCGTTCCGCGAGCCCGACAGCCGATGG 1560  
Db IleThrValIleAlaProHisAlaLeuGlnAenIleValArgGlnProProAspArgTrp 520  
QY 1561 ATGACCAACGCCCGCATGACCCACTATCAAGACCTGCTTCTCAGAGAGAGGCTCACGCTTC 1620  
Db MetThrAenAlaArgMetThrHisTyrglnSerLeuLeuLeuThrGluArgValThrPhe 540  
QY 1621 GCTCCAGCAGCGCTCTCAACCCCTGCCACTCTCTTCTGCTGAAGAGACTGATGAACCAAGTG 1680  
Db AlaProProAlaAlaLeuAenProAlaThrLeuLeuProGluGluThrAspGluProVal 560  
QY 1681 ACTCATGATTCGCCATCAACTATTGATTGAGGAGACTGGGTCGCCAAGGACCTTACAGAC 1740  
Db ThrHisaspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 580  
QY 1741 ATACCGCTGACTGGAGAAGTCTAACTGTTCTACTGACGGAAGCAGCTATGTGGTGAA 1800  
Db IleProLeuThrGlyGluValIleuThrTrpPheThrAspGlySerSerTyrrValValGlu 600  
QY 1801 GGTAAAGAGATCGCTGGGGCGCGTGTGTGACGGGACCCGCCACGATCTGGGGCCAGCAGC 1860  
Db GlyLysArgMetAlaGlyAlaAlaValAlaAspGlyThrArgThrIle\*\*AlaSerSer 620  
QY 1861 CTGCGGAGGAAGAACTTCAGACAAAGGCTGAGCTCATGCGCCCTCAGCAGAGCTTTCGGG 1920  
Db LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
QY 1921 CTGCGGAGGAAGAAATCATATAACATTTATACGACAGCAGGATGCTTTCGACCTGCA 1980  
Db LeuAlaGluGlyLysSerIleAenIleTyrrThrAspSerArgTyrrAlaPheAlaThrAla 660  
QY 1981 CACGTACATGGGCGCATCTATAACAAAGGGGGTGTCTTACTCAGCAGGAGGAGAAATA 2040  
Db HisValHisGlyAlaIleTyrrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
QY 2041 AAGAACAAAGAGGAATCTTAAGCCTATTAGACCGCTGACATTTTACCAAAAGGCTAGCT 2100  
Db LysAenLysAspGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 700  
QY 2101 ATTATACACTGTCTCGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCAAGTG 2160  
Db IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAenGlnMet 720  
QY 2161 GCTGACCGGTTGCCAAGCGGCGAGCCCGGTTGTTAACTTCTGCTTATAATAGAAATG 2220  
Db AlaAspArgIleAlaLysGlnAlaAlaGlnAlaValAenLeuLeuProIleIleGluThr 740  
QY 2221 CCCAAAGCCCGACAGACCCAGACGACGCTACACCTAGAACCTGGCAAGAGATATAAAG 2280  
Db ProLysAlaProGluProArgArgGlnTyrrThrLeuGluAspTrpGlnGluIleLysLys 760  
QY 2281 ATAGACAGGTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGGAATC 2340  
Db IleaspGlnPheSerGluThrProGluGlyThrCysTyrrThrSerTyrrGlyLysGluIle 780  
QY 2341 CTGCCCCCAAGAGAGGTTAGAAATATGTCCAACAGATACATCGTTCACCCACCTAGGA 2400  
Db LeuProHisLysGluGlyLeuGluTyrrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
QY 2401 ACTAAACACCTGACGAGGTTGGTCAGAACATCCCTTATCATGTTCTGAGGCTTACAGGA 2460  
Db ThrLysHisLeu\*\*\*GlnLeuValArgThrSerProTyrrHisValLeuArgLeuProGly 820  
QY 2461 GTGGCTGACTCGGTGGTCAACACTTGTGTCCTGCCCTGCGAGCTGGTAAATGTCTTCTTCC 2520  
Db ValAlaAspSerValValLysHisCysValProCysGlnLeuValAenAlaAenProSer 840  
QY 2521 AGAATGCTCCAGGGAAGAGACTAAGGGGAAGCAACCCAGGCGCTCACTGGGAAGTGGAC 2580

Db 841 LysValProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860  
QY 2581 TTCACTGAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGGTTTGTAGACACC 2640  
Db PheThrLysValLysProAlaLysTyrrGlyAenLysTyrrLeuLeuValPheValAspThr 880  
QY 2641 TTTTCAGATGGGTAGAGGCTTATCTTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 2700  
Db PheSerGlyTrpValGluAlaTyrrProThrLysLysGluThrSerThrValValAlaLys 900  
QY 2701 AAAATACCTGAAGAAATTTTCCAAATTTGGAAATACCTAAGTAAATAGGTCAGACAA 2760  
Db LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAen 920  
QY 2761 GGTCCAGCTTTTGTTCGCCAGGTAAAGTCAAGGACTGCGCCAGATATTGGGGATTGATGG 2820  
Db GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
QY 2821 AAATGCTGATTTGTGCATACAGACCCCAAGCTCAGACAGGCTAGAGAGATGAATAGAAC 2880  
Db LysLeuHisCysAlaTyrrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
QY 2881 ATTAAGAGACCTTACTAAATTGACCCGCGAGACTGGCGTTAATGATTGGATAGCTCTC 2940  
Db IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAenAspTrpIleAlaLeu 980  
QY 2941 CTGCGCTTCTGCTTTTAGGTTAGGAACACCCCTGAGACAGTTGGCTGACCCCTAT 3000  
Db LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrr 1000  
QY 3001 GAATTAATCTACCGGGGACCCCGCTTGGTAGAAATTTGCTGTACATAGTGTCTGAC 3060  
Db GluLeuLeuTyrrGlyGlyProProProLeuValGluIleAlaProValHisSerAlaAsp 1020  
QY 3061 GTGCTGTTTCCAGGCTTTTGTCTTAGGCTCAAGGCACTTGTAGTGGGTGAGACAACGA 3120  
Db ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
QY 3121 GCTTGGAGGCAACTCCGGGAGGCTACTCAGGAGGAGGAGACTTGCAGATCCACATCGT 3180  
Db AlaTrpLysGlnLeuArgGluAlaTyrrSerGlyGluGlyAspLeuGlnValProHisArg 1060  
QY 3181 TTCAAAGTGGAGATTCAGTCTACGTTAGACGCCCGCTGAGGAAACCTCGAGACTCGG 3240  
Db PheGlnValGlyAspSerValTyrrValArgArgHisArgAlaGlyAenLeuGluThrArg 1080  
QY 3241 TGAAGGCGCTTATCTCGTACTTTTGACCCACACCGCTGTGAAGTCGAAGGAATC 3300  
Db TrpLysGlyProTyrrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1100  
QY 3301 TCCACCTGGATCCATCCATCCACGTTAAACCGCGCCACCTCCGATTCGGGTGGA 3360  
Db SerThrTrpIleHisAlaSerHisValLysLeuAlaProProProAspSerGlyTrpArg 1120  
QY 3361 GCCGAAAGACTCAAAATCCCTTAAGCTTCGCTTCCATCGCTGTTCTTACTCTGTC 3420  
Db AlaGluLysThrGluAenProLeuLysLeuArgLeuHisArgLeuValProTyrrSerAsn 1140  
QY 3421 AATAAC 3426  
Db 1141 AsnAsn 1142  
RESULT 14  
Q8Q6U4\_9GAMR PRELIMINARY; PRT; 1142 AA.  
ID Q8Q6U4\_9GAMR  
AC Q8Q6U4;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Poi protein.  
GN Name:pol;  
OS Porcine endogenous retrovirus.

OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
ON NCBI\_taxID=61673;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21851083; PubMed=11861838;  
RX DOI=10.1128/JVI.76.6.2714-2720.2002;  
RA Niebert M., Rogel-Gallard C., Chardon P., Toenjes R.R.;  
RT "Characterization of Chromosomally Assigned Replication-Competent Gamma  
RT Porcine Endogenous Retroviruses Derived from a Large White Pig and  
RT Expression in Human Cells";  
RL J. Virol. 76:2714-2720(2002).  
DR EMBL; AF435967; AAL67856.1; -; Genomic\_DNA.  
DR HSSP; P03355; 1MWL.  
DR MEROPS; A02.020; -.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR PROSITE; PS50879; RNase H; 1.  
SQ SEQUENCE 1142 AA; 127712 MW; B0F55CA742AAC98F CRC64;  
  
Alignment Scores:  
Pred. No.: 0 Length: 1142  
Score: 5466.50 Matches: 1042  
Percent Similarity: 93.4% Conservative: 26  
Best Local Similarity: 91.1% Mismatches: 73  
Query Match: 88.4% Indels: 3  
DB: 2 Gaps: 2  
  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x Q086U4\_9GMR (1-1142)  
  
QY 1 ATGGGTGCCAGGCGCAACACAGTATCCATGGAGTACCCGGAAGAACAGTTGACTTGGGA 60  
DB 1 MetGlyAlaThrGlyGlnGlnThrProThrArgArgThrValAspLeuGly 20  
  
QY 61 GTGGACGGGTAAACCACTGCTTTCTGTCTATACCTAGTGCCTCCAGCACCCCTCTTAGT 120  
DB 21 ValGlyArgValThrGlnSerPheLeuGlyIleProGluCysProGluProLeuLeuGly 40  
  
QY 121 AGAGACTTATTGACCAAGATGGAGCACAAATTTCTTTGAACAGGGAACCAAGAGTG 180  
DB 41 ArgAspLeuLeuThrThrMetGlyArgGlnIlePhePheGluGlnGlyThrProGluVal 60  
  
QY 181 TCTGCAATAACAAACCTATCACTGTGTGACCTCCCAATTAGATGACGAATATCGACTA 240  
DB 61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluThrArgLeu 80  
  
QY 241 TACTCTCCCTAGTAAACCTGATCAAAATATACAAATTCGTGGTGGAAACAGTTCCCCAA 300  
DB 81 PheSerProProValLysLeuAspGlnAsnIleGlnPheGlySerThrGlnPheProGln 100  
  
QY 301 GCCTGGGAGAAACCGCAGGATGGTTGGCAAGCAAGTTCCCCCAAGTTATTTCAA 360  
DB 101 AlaLeuAlaGluProAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 120  
  
QY 361 CTGAAGCGGACGTCACCACTGTGCTAGTCAGACAGTACCCCTTGTAGTAAAGAAGCTCAA 420  
DB 121 LeuLysProSerLeuAlaProValProValArgGlnSerProPheSerLysGluAlaArg 140  
  
QY 421 GAAGGAATTCGGCCGATGTCCAAAGATTAATCCAAAGGGCATCTAGTTCTCTGCCAA 480  
DB 141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleIleValProValGln 160  
  
QY 481 TCTCCCTGGAATACTCCCTGCTACCGGTTAGAAGCCCTGGGACTTAATGACTATCGACCA 540

DB 161 SerProThrAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 180  
QY 541 GTACAGGACTTGAGAGAGGTCAATAAACGGGTGAGGATATACACCCCAACAGTCCCGAAC 600  
DB 181 ValGlnAspPheGluArgGlyGlnLysArgValGlnAspIleHisProThrValProAsn 200  
QY 601 CCTTATAACCTCTTGTGTCTCCACCCCAACGGAGCTGTATACAGTATTGGACTTA 660  
DB 201 ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTyrThrValLeuAspLeu 220  
QY 661 AAGGATCCCTTCTTCTGCTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTTCGAA 720  
DB 221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240  
QY 721 TGGAGATCCAGTACGGGAAGAACCCGGGAGCTCACTGGACCCGACTGCCCAAGGG 780  
DB 241 TrpArgAspProGlyAlaGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260  
QY 781 TTCAGAACCTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCCACTTCAGG 840  
DB 261 PheLysAsnPheProThrIlePheAspGlnAlaLeuHisArgAspLeuAlaAsnPheArg 280  
QY 841 ATCCCAACACCTCAGGTGACCCCTCTCCAGTACGTGGATGACCTGCTTCTGGCGGAGCC 900  
DB 281 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuValLeuAlaAla 300  
QY 901 ACCAAACAGGACTCTTAGAAGGACCAAGCAGCTACTGTGGAATTTGTCTGACTTAGGC 960  
DB 301 ThrLysGlnAspCysLeuGlnArgProLysGlyLeuValGluLeuSerAspLeuGly 320  
QY 961 TACAGAGCCTCTGCTAAGAAGGCCAGATTTTGACGAGAGAGAGTAAACATCTCTGGGTAC 1020  
DB 321 AspArgAlaPheGlyTyrLysAlaHisIleCysProThrGluValThrTyrLeuGlyTyr 340  
QY 1021 AGTTTGGGGACCGGCGAGCGATGCTGACGAGGACGACGGAATAAAGTGTAGTCCAGATA 1080  
DB 341 ArgLeuArgGlyArgHisArgTyrLeuThrGluAlaProGlnThrValValGlnIle 360  
QY 1081 CCGGCCCCCAACACAGCCAAACAAATGAGAGAGTTTGGGAGACGTGGATTTTCGAGA 1140  
DB 361 ProGlyProThrProAlaLysGlnValArgGluPheLeuGlyThrValGlyPheCysArg 380  
QY 1141 CTGTGGATCCCGGGTTTGGACCTTAGACCCCACTCTACCCGCTAACCAAGAAAAA 1200  
DB 381 LeuThrIleProGlyPheAlaThrLeuProAlaProLeuTyrProLeuProLysGlnLys 400  
QY 1201 GGGGAATTTCTCTGGGCTCTGAGCACCAAGGCATTTGATGCTATCAAAAAGGCCCTG 1260  
DB 401 GlyGluPheSerTrpAlaLeuGlnHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420  
QY 1261 CTGAGCGACCTGCTCTGGCCCTCCCTGACGTAAACCCCTTTACCTTTATGTGTGAT 1320  
DB 421 LeuSerAlaProAlaLeuAlaLeuProAspValThrLys-----ThrLeuTyrValAsp 438  
QY 1321 GAGGTAAAGGAGTAGCCCGGGAGTTTAAACCAACCCCTAGGACCATGAGAGACCT 1380  
DB 438 LeuSerAlaProAlaLeuAlaLeuProAspValThrLys-----ThrLeuTyrValAsp 458  
QY 439 GluArgLysGlyValAlaAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 458  
QY 1381 GTCGCTACCTGTCTCAAGAAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCTGTGAAG 1440  
DB 459 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTyrProIleCysLeuLys 478  
QY 1441 GCTATCGAGCTGTGGCCATCTGGTCAAGACCTGACCAATTCAGCTTTGGGACAGAT 1500  
DB 479 AlaIleAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 498  
QY 1501 ATAACTGTAATAGCCCCCATGTCATTGGAGAACATCGTTCCGCGACGCCCCAGACCGATGG 1560  
DB 499 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 518  
QY 1561 ATGACCAACGCCCATGACCCCACTATCAAAAGCCTGCTTCTCACAGAGGGGTACAGTTC 1620

Db 519 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuLeuThrGluArgValThrPhe 538  
QY 1621 GCTCCACGCGCTCTCAACCCCTGCCACTCTCTTCCTGCTGAAGAGACTGATCAACACGATG 1680  
Db 539 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluLeuThrAspGluProLeu 558  
QY 1681 ACTCATGATTGCCATCACTATTGATTGAGGAGACTGGGGTCCGCAAGGACTTACAGAC 1740  
Db 559 ThrHisAspCysHisGlnLeuLeuLeuGluThrGlyValArgLysAspLeuThrAsp 578  
QY 1741 ATACCGCTGACTGGGAGAAGTCTAACCTGGTTCTACTGACGCAAGCAGCTATGTGGTGA 1800  
Db 579 IleProLeuThrGlyGluProValThrTrpPheThrAspGlySerSerTrpLeuValGlu 598  
QY 1801 GGTAAAGAGATGGCTGGCGCGCGTGTGGACGGGACCCGACGATCTGGGGCCAGCAGC 1860  
Db 599 GlyAsnLysMetAlaGlyAlaAlaValValAspArgThrProThrIleTrpGlyThrAsn 618  
QY 1861 CTGGCGGAAGNACTTCAGCACAAAGGCTGAGCTCATGCGCCCTCACGCAAGCTTTGGG 1920  
Db 619 LeuProGluArgThrSerSerGlnLysGlyLeuLeuIleGlyLeuMetGlnAlaPheArg 638  
QY 1921 CTGGCCGAAGGAAATCCATAAATTTATACGGACAGCAGTATGCTTTGGCACTGCA 1980  
Db 639 LeuGlyGlnGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 658  
QY 1981 CACGTACATGGGCGCATCTATAAACAAGGGGGTGTCTTACTCAGCAGGAGGAGAAATA 2040  
Db 659 HisValHisGlyAlaIleTyrThrGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 678  
QY 2041 AAGAACAAAGAGAAATTTCTAAGCTATTAGACCGGTACATTTACCAAAAGGCTAGCT 2100  
Db 679 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 698  
QY 2101 ATTATACATCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATG 2160  
Db 699 IleIleHisCysProGlyHisGlnLysAlaLysAspProIleSerArgLysAsnGlnMet 718  
QY 2161 GCTACCGGGTGGCAAGCAGGAGCCAGGGTGTAACTTCTGCTCTATAATAAGAAATG 2220  
Db 719 AlaAspArgValAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProMetIleGluThr 738  
QY 2221 CCCAAGCCCCAGAACCCAGACGACGATCACCTTAGAAGACTGGCAAGAGATAAAAG 2280  
Db 739 ProLysAlaProGluProGlyArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 758  
QY 2281 ATACACCATGTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGATC 2340  
Db 759 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 778  
QY 2341 CTGCCCCAACAAAGGGTTAGAAATATGTCCACAGATACATCTCTAAACCCACCTAGGA 2400  
Db 779 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 798  
QY 2401 ACTAAACACTGACGAGTGGTCAGAACATCCCTTATCATGTTCTGAGGCTACCAAGGA 2460  
Db 799 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 818  
QY 2461 GTGGCTGACTCGGGTCAACATTTGTGCTCCGCGGCTGGTGAATGCTTAATCCTTCC 2520  
Db 819 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 838  
QY 2521 AGAATGCTCCAGGGAAGAGACTAAGGGGAAGCCACCGCGCTCACTGGGAAGTGCAC 2580  
Db 839 ArgIleProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 858  
QY 2581 TTCACTGAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGTTTGTGTAGACACC 2640  
Db 859 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 878  
QY 2641 TTTTCAGATGGGTAGAGGCTTATCTCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA 2700  
Db 879 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 898

QY 2701 AAAATACTGGAAGAAATTTTTCAGGATTTTGAATACCTAAAGTAATAGGCTCAGACAAT 2760  
Db 899 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 918  
QY 2761 GGTCCAGCTTTTGTGGCCAGGTAAAGTACAGGACTGCGCAAGATATTGGGAGTATTGG 2820  
Db 919 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 938  
QY 2821 AAACCTGCATGTGTCATACAGACCCCAAGCTCAGCAGCAGGTAGAGAGGATCAATAGACC 2880  
Db 939 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 958  
QY 2881 ATTAAGAGAGACCTTACTAAATTCACCGCGAGAGACTGCGCTTAATGATTGGATAGCTCTC 2940  
Db 959 IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeu 978  
QY 2941 CTGCCCCTTTGTCTTTTAGGTTAGGAACAACCCCTGGACAGTTTGGGCTGACCCCTTAT 3000  
Db 979 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 998  
QY 3001 GAATTACTCTACGGGGGACCCCGCCCATTTGGTAGAAATTTGCTTCTGTACATAGTCTGAC 3060  
Db 999 GluLeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisSerAlaAsp 1018  
QY 3061 GTGCTGCTTTCCAGACCTTTTCTCTAGGCTCAAGGCACCTTGAGTGGGTGAGACAACGA 3120  
Db 1019 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1038  
QY 3121 CGGTGAGGCAACTCCGGGAGGCTTACTCAGGAGGAGAGACTTTCAGATCCCACTCGT 3180  
Db 1039 AlaTrpLysGlnLeuArgGluAlaTyrSer---GlyGlyAspLeuGlnValProHisArg 1057  
QY 3181 TTCCAAGTGGGAGATTCACTAGTGTAGACGCCCGCTCAGGAAACCTCGAGACTCGG 3240  
Db 1058 PheGlnValGlyAspSerValTyrValArgHisArgAlaGlyAsnLeuGluThrArg 1077  
QY 3241 TGAAGGCGCTTATCTCTGCTACTTTTCCACACACCAACCGGTGTGAAGTCAAGAGGAATC 3300  
Db 1078 TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1097  
QY 3301 TCCACTGGATCCATGATCCACGTTAAACCGCGCCACCTCCCGGATCGGGGTGGA 3360  
Db 1098 ProThrTrpIleHisAlaPheHisValLysProAlaProSerAspSerGlyTrpLys 1117  
QY 3361 GCCGMAAGACTGAAATCCCTTAGCTTCGCTCCATCGCTGGTTCCTTACTCTGTC 3420  
Db 1118 AlGlnLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1137  
QY 3421 AATAACCTCTCA 3432  
Db 1138 AsnAsnSerSer 1141

## RESULT 15

Q8Q6U7\_9GAMR  
ID Q8Q6U7\_9GAMR PRELIMINARY; PRT; 1139 AA.  
AC Q8Q6U7;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Pol protein.  
GN Name=pol;  
OS Porcine endogenous retrovirus.  
OC Viruses; Retrovirdae; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21851083; PubMed=11861838;  
RX DOI=10.1126/JVI.76.6.2714-2720.2002;  
RA Niebert M., Rogel-gallard C., Chardon P., Toenjes R.R.;  
RT "Characterization of Chromosomally Assigned Replication-Competent Gamma  
Porcine Endogenous Retroviruses Derived from a Large White Pig and



```

RT Expression in Human Cells.;
RL J. Virol. 76:2714-2720(2002).
DR EMBL; AF435966; AAL87853.1; -, Genomic_DNA.
DR HSSP; P03355; IMML.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006313; F:DNA transposition; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; RVT; 1; 1.
DR PROSITE; PS50879; RNase H; 1.
SO SEQUENCE 1139 AA; 127325 MW; 3B63C42301C04PAC CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1139
Score: 5183.50 Matches: 1054
Percent Similarity: 94.0% Conservative: 19
Best Local Similarity: 92.3% Mismatches: 52
Query Match: 87.1% Indels: 17
DB: 2 Gaps: 13

US-10-723-552-3_COPY_2307_5741 (1-3435) x Q8Q6U7_9GAMR (1-1139)
QY 41 GAAGAAACAGTTGACTGGGAGTGGGACGGGTAAACCACTGCTTTCTGTGTCATACCTG--- 97
DB 4 GluAsnGlnLeuThrTrpGluLeuAspGlyGluProThrGlyPheTrpValIleLeuLys 23
QY 98 AGTGCCAGCACCCCTCTTAGTAGAGACTTATTGACCAAGATGGAGTGGGACCAAAATTTTC-- 155
DB 24 Cys-ProSerThrLeuLeuArgValGlnThrTyTrpAspGlnAspGlyLysProLysPheLe 43
QY 156 -TTTTGAACAGGAAACC---AGAAAGTGTCTGCAAAATPAACAAACCTATCATCTGTGTGA 211
DB 43 uPheGluProArgGluThrGlnLysCysLeu-GlnAsnThrAsnProIleThrValLeuT 63
QY 212 CCCTCCAAATAGATGACGAATATCGACTATACTCTCCCTAGTAAAGCTCATCAAAATA 271
DB 63 hrLeuGlnLeu---AspGluTyArgLeuTySerProLeuValLysProAspGlnAsnI 82
QY 272 TACAATCTGTTGGACAGTTTCCCAAGCTCGGCGAAGAACCGAGGGATGGTTTGG 331
DB 82 leGlnPheTrpGluGlnPheProLysAlaTrpAlaGluThrAlaGlyMetGlyLeuA 102
QY 332 CAAAGCAAGTTCCTCCCAACAGTTATTCAACTGAAGCCAGTGCCACACCACTGTCAGTCA 391
DB 102 laLysGlnValProProGlnValIleGlnLeuLysAlaSerAlaAlaProValSerValA 122
QY 392 GACAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTCGGCCCGCATGTCCAAAGATTAA 451
DB 122 rgGlnTyLeuLeuSerLysGluAlaArgGluGlyIleGlyProHisValGlnArgLeuI 142
QY 452 TCCAAACGGGCATCTAGTCTGTCGCAATCTCCCTGGAATACTCCCTGCTACCGGTTA 511
DB 142 leGlnGlnGlyIleLeuValProValGlnSerProTrpAsnThrProLeuLeuProValA 162
QY 512 GAAAGCCTGGGCAATTAATGACTATCGACCAAGTACAGACTTGAGAGAGTCAATAAACGGG 571
DB 162 rgLysProGlyThrAsnAspTyArgProValGlnAspLeuArgGluValAsnLysArgV 182
QY 572 TGCAGGATATACACCAACAGTCCCGAACCCCTTAACTCTTGTGTCTCTCCACCCC 631
DB 182 alGlnAspIleHisProThrValProAsnProTyAsnLeuLeuCysAlaLeuProProG 202
QY 632 AACGAGCTGGTATACAGTATTGGACTTTAAAGGATGCCTTCTTCTGCTGAGATTACACC 691

```

---

```

202 lnArgSerTrpTyThrValLeuAspLeuLysAspAlaPhePheCysLeuArgLeuHisP 222
692 CCCTAGCCCAACCACTTTTTCCTTCCAAATGGAGAGATCCAGGTACGGGAGAACCCGGGC 751
222 roThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyAlaGlyArgThrGlyG 242
752 AGCTCACCTGGACCCGACTGCCCAAGSGTTCAAGAACTCCCGACCATCTTTTTCACGAAG 811
242 lnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrIlePheAspGluA 262
812 CCCTACACAGAGACTGCGCCAACTTCAGGATCCAAACACCTCTCAGGTGACCCCTCTCCAGT 871
262 laLeuHisArgAspLeuAlaAsnPheArgIleGlnHisProGlnValThrLeuLeuGlnT 282
872 ACGTGGATGACCTGCTCTGCGGGAGCCACCAACAGGACTGCTTGAAGAGCCACGAGG 931
282 yrValAspAspLeuLeuAlaGlyAlaThrLysGlnAspCysLeuGluGly----- 299
932 CACTACTGCTGGAATTCCTGACCTAGCTACAGAGCTCTGCTAAGAAAGCCACAGATTT 991
300 --LeuLeuLeuGluLeuPheAspLeuGlyTyArgAlaSerAlaLysLysAlaGlnIleC 319
992 GCAGGAGAGAGTTAAACATCTTGGGGTACAGTTTGG---CGGGACGGGCAGCGA---TGGC 1045
319 ysArgAGGluAlaThrGlnLeuGlyValGlnValCysGlyAlaGlyGlnSerAspTrpL 339
1046 TGACG---GAGGCGACGGAAGAAACTGTA---GTCCAGATACCGGCCCCCAACACGACCA 1099
339 euThrGlyLysAlaArgLysLysThrValGlnProLysIleGlyProProThrThrAlaL 359
1100 AACAA---ATCAGAGAGTTTTTGGGGACA---GCTGGATTTTGCAGACTGGGATCCCGG 1153
359 ysGlnValValArgGluPheGlyAlaGlnValGlyPheCysArgLeuTrpIleProG 379
1154 GGTTTGGGACCTTAGCAGCCCCACTCTACCGCTTAACCAAGAAAGGCGCTGCTGAGCGCACCTG 1213
379 lyPheAlaThrLeuAlaAlaProLeuTyProLeuThrLysGluLysGlyGluPheSerT 399
1214 GGGCTCTCTGAGCACCAGAGGCATTTGATGCTATCAAAAGGCGCTGCTGAGCGCACCTG 1273
399 rpAlaLeuGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeuSerSerAlaProA 419
1274 CTCTGGCCCTCCCTGAGCTAAACCTTTACCTTTATGCTGATGAGCGTAAAGGAG 1333
419 laLeuAlaLeuProAspValLeuLysProPheThrLeuTyValAspGluArgLysGlyV 439
1334 TAGCCCGGGAGTTTAAACCCAAACCTTAGCACCATGGAGAGAGCCTGTGCGCTACCTGT 1393
439 alAlaArgGlyValLeuThrGlnIleLeuGlyProTrpArgProValAlaTyLeuS 459
1394 CAAAGAGCTCGATCTCTGTAGCCAGTGGTGGCCCATATGCTGAGGCTATCGCAGCTG 1453
459 erLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLysAlaIleAlaAlaV 479
1454 TGGCCATACCTGGTCAAGAGCGCTCACAAATTTGACTTTGGGACAGAAATATACTGTAATAG 1513
479 alAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsnIleThrValIleA 499
1514 CCCCCCATGCTATGGAGAACATCGTTGCGGACGCCCCAGACCCGATGGATGATGCCAACGCC 1573
499 laProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrpMetThrAsnAlaA 519
1574 GCATGACCCACTATCAAGCCTGCTTCTCAGAGAGGGTTCAGTTCGCTCCACGAGCGG 1633
519 rgMetThrHisTyGlnSerLeuLeuLeuThrGluArgValThrPheAlaProProAlaA 539
1634 CTCTCAACCCCTGCCACTCTTCTGCTGAAGAGACTGATGAACCACTGACTCATGATTGCC 1693
539 laLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProValThrHisaspCysH 559
1694 ATCAACTATTGATTGAGGAGACTGGGTCCGCAAGGACCTTTACAGACATA---CGCTGA 1750

```

Db 559 is---LeuLeuIleGluThrGlyValArgLysAspLeuThrAspIleProProLeuT 578  
Qy 1751 CTGAGAGAGTCTAACCTGGTTCTACTGACGGAAGCAGCTATGTGTGGAGGTGAAGGA 1810  
Db : : : : :  
Db 578 hrGlyLysMetLeuThrTrpPheThrAspGlySerSerTyrValValGluGlyLysSerM 598  
Qy 1811 TGGCTGGGGCGGCGTGTGACGGGACCCGACGATCTGGCCAGCGACCTGCCGAAG 1870  
Db 598 etAlaGlyProProValValThrGlyThrArgThrIleTrpAlaSerSerLeuProGluG 618  
Qy 1871 GAACTTCAGCAAAAGGCTCAGCTCATGGCCCTCAGCAAGCTTTGCGGCTGCCGAAG 1930  
Db 618 lyThrSerAlaGlnLysAlaGlnLeuMetAlaLeuThrGlnAlaLeuArgLeuAlaGluG 638  
Qy 1931 GGAATCCATAAACATTATACGACAGCAGGTATGCCCTTTGCGACTGCACACGTACATG 1990  
Db 638 lylsSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAlaHisValHisG 658  
Qy 1991 GGGCCATCTATAACAAAGGGGTGCTTACTCAGCAGGGAGGGAATAAAGAACAAAG 2050  
Db 658 lyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIleLysThrLysG 678  
Qy 2051 AGGAAATCTAAGCCTATTAGAAGCGGTACATTTACCANAAGGCTAGCTATTATACACT 2110  
Db 678 luGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAlaIleIleHisC 698  
Qy 2111 GTCCTGGACATCAGAAAGCTAAAGATCTCATATCAGAGGAAACAGATGGCTACCGGG 2170  
Db 698 ysProGlyHisGlnLysAlaLysAspProIleSerArgLysGlnMetAlaAspArgV 718  
Qy 2171 TTGCCAAGCAGGCGACCCAGGGTGTAACTTCTGCTTAAATAAGAAATGCCAAAGCCC 2230  
Db 718 alAlaLysGlnAlaGlnGlyValAsnLeuLeuProMetIleGluThrProLysAlaP 738  
Qy 2231 CAGAAACCCAGCAGCAGTACACCTCAGAGCTGCGAAGAGATAAAGATAGACCCAGT 2290  
Db 738 roGluProGlyArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLysIleAspGlnP 758  
Qy 2291 TCTCTGAGACTCCGGAAGGACCTGCTATATCCTCAGATGGGAAGAAATCCTGCCCCACA 2350  
Db 758 heSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIleLeuProHisL 778  
Qy 2351 AAGAGGGTTAGAAATATGTCACAGATACATCTCTAACCCACTAGGAACATAACACC 2410  
Db : : : : :  
Db 778 ysGluGlyLeuLysTyrValGlnGlnIleHisArgLeuThrHisLeuGlyThrLysHisL 798  
Qy 2411 TGACGACAGTTGGTCAGAACATCCCCTTATCATGTTCTCAGGCTACACGAGGTGGCTCACT 2470  
Db 798 euGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGlyValAlaAsps 818  
Qy 2471 CGGTGGTCAAAACATTGTGTGCCCTGCCAGCTGGTTAATGCTTAATCTTCCAGAAATGCTC 2530  
Db 818 erValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSerArgIleProp 838  
Qy 2531 CAGGGAAGAGACTAAGGGGAAGCACCAGCGCTCACTGGGAAGTGGACTTCACTGAGG 2590  
Db 838 roGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAspPheThrGluV 858  
Qy 2591 TAAAGCCGGCTTAATACGGAACAAATACCTATTTGGTTTTGTAGACACCTTTTCAGGAT 2650  
Db 858 alLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThrPheSerGlyT 878  
Qy 2651 GGGTAGAGGCTTATCTTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAAAAAAATCTGG 2710  
Db 878 rpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLysLysIleLeuG 898  
Qy 2711 AAGAAATTTTCCAAAGATTTGGAATACCTAAGGTAAATAGGGTTCAGACAAATGGTCCAGCTT 2770  
Db 898 luGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsnGlyProAlaP 918  
Qy 2771 TTGTTGCCAGGTAAGTCAGGACTGCCAAGATATTGGGGATTGATTGGAATGCAATT 2830  
Db 918 heValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrpLysLeuHisC 938

Qy 2831 GTGCATACAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATAGAACCACTTAAAGAGA 2890  
Db : : : : :  
Db 938 ysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluT 958  
Qy 2891 CCCTTACTAAATTCAGCCGAGACTGGCGTTAATGATTTGGATAGTCTCTCTGCCCTTTG 2950  
Db 958 hrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTrpIleAlaLeuLeuProPheV 978  
Qy 2951 TGCCTTTTTCAGGTTTTCAGCAACCCCTCGACAGTTTGGCGTGGACCCCTATCAATTAATCT 3010  
Db 978 alLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyrGlnLeuLeuT 998  
Qy 3011 ACGGGGACCCCCCACTTGTGATAATTGCTTCTGTACATAGTGTGCTGACGTGCTGCTTT 3070  
Db 998 yrGlyGlyProProProLeuValGluIleAlaSerValHisSerAlaAspValLeuLeuS 1018  
Qy 3071 CCCAGCCTTTGTTCTTCTAGGCTCAAGGCATCTGAGTGGGTGAGACAACGAGCGTGGAGGC 3130  
Db 1018 erGlnProLeuPheSerArgLysAlaLeuGluTrpValArgGlnArgAlaTrpLysG 1038  
Qy 3131 AACTCCGGGAGGCTACTCAGAGGAGGAGACTTGCAGATCCCAATCGTTTCCAAGTGG 3190  
Db 1038 lnLeuArgGluAlaTyrSerGlyGluGlyAspLeuGlnValProHisArgPheGlnValG 1058  
Qy 3191 GAGATTTCAGTCTACTGTAGACGCCCGCTGAGGAAACCTCGAGACTCGGTGGAGAGGCC 3250  
Db 1058 lyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArgTrpLysGlyP 1078  
Qy 3251 CTTATCTCGTACTTTTTCACCAACCGCTGTGAAAGTCGAAAGGAATCTCCACCTGGA 3310  
Db 1078 roTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIleSerThrTrpI 1098  
Qy 3311 TCCATGCATCCCGGTTAAACCGCGCCACTCCCGATTCCGGGTGGAAAGCCGAAAGA 3370  
Db 1098 leHisAlaSerHisValLysProAlaProProAspSerGlyTrpLysAlaGluLysT 1118  
Qy 3371 CTGAAATCCCTTAAGCTTTCGCTTCCATCGCTGGTTCCCTTACTCTGTCAATTAACCTCT 3430  
Db 1118 hrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerValAsnAsnSerS 1138  
Qy 3431 CA 3432  
Db : :  
Db 1138 er 1138

Search completed: February 14, 2006, 14:50:11  
Job time : 978.589 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 12:52:14 ; Search time 91.3916 Seconds  
(without alignments)  
3302.856 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_2307\_5741

Perfect score: 6183

Sequence: 1 ATGGGTGCCAGGCAACA.....CTGTCAATACCTTCAGAC 3435

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB.spool/US10723552/runat\_14022006\_125139\_12833/app\_query.fasta\_1  
-DB=A Geneseq -QFMT=fastan -SURFIX=rag -MINMATCH=0.1 -LOOCL=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCLALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p  
-USER=US10723552 @CGN 1 1 734 @runat\_14022006\_125139\_12833 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6038	97.7	1145	2 AAW32097	Miniature
2	6038	97.7	1145	4 AAB73286	Retroviral
3	6038	97.7	1145	8 ADS73447	Swine ret
4	5851.5	94.6	1194	2 AAW39272	Porcine r
5	4146.5	67.1	1203	2 AAW81572	Mus dunni
6	3818	61.7	1196	2 AAR75189	Osteoindu
7	3816.5	61.7	1197	4 AAB49468	Canine re
8	3808.5	61.6	1199	7 ADH76470	Murine le
9	3797.5	61.4	1199	3 AAB12994	MLV rever

10	3797.5	61.4	1737	3 AAB10044	MMLV gag-
11	3781	61.2	1224	2 AAY17947	MoMLV pol
12	3696	59.8	1784	2 AAR94427	FelV F6A
13	3686	59.6	1784	2 AAR05898	Gene prod
14	3600	58.2	1079	2 AAY03163	MuLV reve
15	3090	50.0	1193	5 ABB79876	Spleen ne
16	2827	45.7	540	8 ADS73451	Swine ret
17	2491	40.3	672	3 AAB12457	Moloney m
18	2487.5	40.2	716	5 AAU74989	Moloney m
19	2487.5	40.2	716	8 ADN48501	M-MuLV re
20	2487	40.2	671	9 AEB46280	Murine le
21	2486	40.2	671	9 AEB46278	Murine le
22	2485.5	40.2	683	5 AAU74994	Moloney m
23	2482.5	40.2	683	5 AAU75012	Moloney m
24	2480.5	40.1	683	5 AAU75010	Moloney m
25	2480.5	40.1	683	5 AAU75017	Moloney m
26	2480.5	40.1	683	5 AAU74995	Moloney m
27	2480.5	40.1	683	5 AAU75009	Moloney m
28	2480.5	40.1	683	5 AAU75014	Moloney m
29	2480.5	40.1	683	5 AAU75011	Moloney m
30	2479.5	40.1	683	5 AAU74992	Moloney m
31	2479.5	40.1	683	5 AAU74997	Moloney m
32	2478.5	40.1	683	5 AAU75004	Moloney m
33	2478.5	40.1	683	5 AAU74990	Moloney m
34	2477.5	40.1	683	5 AAU75013	Moloney m
35	2476.5	40.1	683	5 AAU74991	Moloney m
36	2476.5	40.1	683	5 AAU74998	Moloney m
37	2476.5	40.1	683	5 AAU75008	Moloney m
38	2476.5	40.1	683	5 AAU75020	Moloney m
39	2476.5	40.1	683	5 AAU74999	Moloney m
40	2418.5	39.1	1183	8 ADW50751	Human pro
41	2221.5	35.9	1069	4 AAB73284	Defective
42	2051	33.2	401	2 AAW32092	Porcine r
43	2051	33.2	401	4 AAB73283	Defective
44	2051	33.2	401	8 ADS73442	Swine ret
45	2013	32.6	471	5 ABP47783	Protein #

ALIGNMENTS

RESULT 1

AAW32097

ID AAW32097 standard; protein; 1145 AA.

XX

AC AAW32097;

XX

DT 27-AUG-2003 (revised)

DT 09-FEB-1998 (first entry)

XX

DE Miniature swine retrovirus POL protein.

XX

Retrovirus; porcine; POL protein; xenotransplantation; infectious; provirus; organ transplant; donor; activated virus; PCR.

XX

OS Pig endogenous retrovirus.

XX

XX Key Location/Qualifiers

FT Protein

FT 1..1145 /label= POL\_protein

XX

XX WO9721836-A1.

XX

XX 19-JUN-1997.

XX

XX 13-DEC-1996; 96WO-US019680.

XX

XX 14-DEC-1995; 95US-00572645.

XX

XX (GEO ) GEN HOSPITAL CORP.

XX

XX Fishman JA;

XX

XX WPI; 1997-332804/30.

DR

DR N-PSDB; AAT74884.  
 XX New nucleic acid from porcine retroviruses - used for detecting viruses  
 PT in transplant or other tissue and for assessing risk of transmitting  
 PT infection to graft recipient.  
 XX  
 XX  
 PS Claim 22; Fig 3; 128pp; English.  
 XX  
 CC This is a porcine retrovirus from miniature swine containing the coding  
 CC region for a putative viral POU protein. This sequence and PCR fragments  
 CC generated from the sequence (see AAT74812-T74882) could be used to screen  
 CC organs for porcine retroviruses prior to xenotransplantation.  
 CC Transplantation can increase the likelihood of retroviral activation if  
 CC intact and infectious proviruses are present. The porcine retroviral  
 CC sequence can be used to generate probes to determine the level (e.g. copy  
 CC number) of intact (i.e. potentially replicating) porcine provirus  
 CC sequences in a strain of xenograft transplantation donors. It can be used  
 CC to detect mutations, genetic lesions or viral recombinants and also to  
 CC determine the histological localisation of activated retrovirus. Using  
 CC Polymerase Chain Reaction DNA Quantitation (PDQ) on blood mononuclear  
 CC cells, infectivity titration and susceptibility testing can be performed.  
 CC Ultimately animal donors without intact porcine retroviral sequences or a  
 CC lower copy number of viral elements could be selected. (Updated on 27-AUG  
 CC -2003 to correct OS field.)  
 XX

SQ Sequence 1145 AA;

#### Alignment Scores:

Pred. No.: 0 Length: 1145  
 Score: 6038.00 Matches: 1145  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 97.7% Indels: 0  
 DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x AAW32097 (1-1145)

QY 1 ATGGGTGCACGGGACACACAGTATCCATGGACTACCCGAGACAGTGTGCTGGGA 60  
 DB 1 MetGlyAlaThrGlyGlnGlnGlnTyrProThrThrArgAthrValAspLeuGly 20  
 QY 61 GTGGGACGGGTAAACCCACTCGTTCTGGTGTATACCTGAGTGCACAGCACCCCTCTTAGGT 120  
 DB 21 ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 40  
 QY 121 AGAGACTATTGACCAAGATGGGAGCAAAATTTCTTTTGAACCAAGGAAACCAAGAGTG 180  
 DB 41 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyLysProGluVal 60  
 QY 181 TCTGCAATACAAACCTTACCTGCTGTGTGACCTCCCAATTAGATGACGAATATCGACTA 240  
 DB 61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu 80  
 QY 241 TACTCTCCCTAGTAAAGCCTGTATCAAAATATACAAATTCGTGTTGGAAACAGTTTCCCCAA 300  
 DB 81 TyrSerProLeuValLysProAspGlnAsnIleGlnPheTrpLeuGluGlnPheProGln 100  
 QY 301 GCCTGGGAGAAACCGCAGGATGGTTTGGCAAGCAAGTTCCTCCCAAGATTTATTCAA 360  
 DB 101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 120  
 QY 361 CTGAAGCCAGTGCACACCACTGTCTAGTCCAGACAGTACCCCTTGAGTAAGACAGCTCAA 420  
 DB 121 LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerLysGluAlaGln 140  
 QY 421 GAAGGAATTCGGCCCATGTCTCAAGATTAATCAACAGGGCATCTAGTTCTCTGTCCAA 480  
 DB 141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 160  
 QY 481 TCTCCCTCGAATACTCCCTGCTACCGGTTAGAAAGCCTGGGACTTAATGACTATCGACCA 540  
 DB 161 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 180

QY 541 GTACAGGACTTGAGAGAGTCAATAAACCGGTGAGGATATACACCAACAGTCCCGAAC 600  
 DB 181 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200  
 QY 601 CCTTATTAACCTCTTTGTGTCTCTCCACCCCAACCGAGCTGGTATACAGTATTGGACTTA 660  
 DB 201 ProTyrAsnLeuLeuCysAlaLeuProProGlnArgSerTrpTyrThrValLeuAspLeu 220  
 QY 661 AAGGATCGCTCTTCTGCTGAGATTACACCCACTAGCCACCAACCTTTTGCCTCCAA 720  
 DB 221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240  
 QY 721 TGGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGACCCGACTGCCCAAGGG 780  
 DB 241 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260  
 QY 781 TTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGG 840  
 DB 261 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280  
 QY 841 ATCCAAACCTCAGGTGACCTCTCCAGTACCTGAGTACCTGCTTCTGCGGGAGCC 900  
 DB 281 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla 300  
 QY 901 ACCAAACAGGACTGCTTAGAAGCACGAAGGCACCTACTGCTGGAATTGTCTGACCTAGGC 960  
 DB 301 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuLeuSerAspLeuGly 320  
 QY 961 TACAGAGCCTCTGCTAAGAGGCCAGATTTGACGAGAGAGAGTAACATCTTGGGGTAC 1020  
 DB 321 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 340  
 QY 1021 AGTTTGGGGACGGCGAGCGATGCTGACGAGCAGCAGGAAGAAACTGTAGTCCAGATA 1080  
 DB 341 SerLeuArgAspGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 360  
 QY 1081 CCGGCCCAACACACAGCCAAACAAATGAGAGATTTTGGGGACAGCTGGGATTTGCAGA 1140  
 DB 361 ProAlaProThrThrAlaLysGlnMetArgGluPheLeuGlyThrAlaGlyPheCysArg 380  
 QY 1141 CTGTGATTCGCGGGTTTGGACCTTACGACCCCTTACGACCCCTTACCCCTTAACCAAGAAAA 1200  
 DB 381 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLys 400  
 QY 1201 GGGGAATTCCTCGGCTCCTGACACACAGAGCATTTGATGCTATCAAAAGGCCCTG 1260  
 DB 401 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420  
 QY 1261 CTGAGCGCACCTGCTCTGGCCCTCCCTGACGTAACTAAACCTTTTACCTTTATGTGGAT 1320  
 DB 421 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 440  
 QY 1321 GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCAACCCCTTAGGACCATGGAGAGACT 1380  
 DB 441 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 460  
 QY 1381 GTCCCTACCTGTCAAAGAAGCTCGATCTGTAGCAGTGGTGGCCCATATGCTGAG 1440  
 DB 461 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProIleCysLeuLys 480  
 QY 1441 GCTATCGCAGCTGTGGCCCATCTGCTCAAGACGCTGACAAATGACTTTTGGGACAGAA 1500  
 DB 481 AlaIleAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 500  
 QY 1501 ATAACTGTAATAGCCCCCATGCTGATTTGGAGAACATCGTTCCGCGAGCCCCCAGACCGATGG 1560  
 DB 501 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 520  
 QY 1561 ATGACCAACGCCCGCATGACCCACTATCAAGCCTGCTTCTCACAGAGAGGGTCACGGTTC 1620  
 DB 521 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 540  
 QY 1621 GCTCCACCGCCGCTCTCAACCCCTGCCACTCTTCTGCTGGAAGAGACTGATGAACCACTG 1680

Db 541 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 560  
QY 1681 ACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGACCTTACAGAC 1740  
Db 561 ThrHisaspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 580  
QY 1741 ATACCGTGACTGGGAAGTGTCTAACCTGGTTCACTGACGGAAGCAGCTATGTGGTGA 1800  
Db 581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerFyrValValGlu 600  
QY 1801 GGTAAAGAGTGGTGGGGCGGTGGTGGACGGGACCCACAGATCTGGCCAGCAGC 1860  
Db 601 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 620  
QY 1861 CTGCCGGAAGGAACCTTCAGCACAAAGGCTCAGGCTCATGGCCCTCAGCAAGCTTTGGG 1920  
Db 621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
QY 1921 CTGCCGGAAGGAATCATTAACATTTATACGACAGCAGGTATGCCTTTGGCAGCTGCA 1980  
Db 641 LeuAlaGluGlyLysSerIleAsnIleFyrThrAspSerArgTyrAlaPheAlaThrAla 660  
QY 1981 CACGTACATGGGCGCATCTATAACAAAGGGGTGCTTACCTCAGCAGGAGGAAATA 2040  
Db 661 HisValHisGlyAlaIleFyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
QY 2041 AAGAAACAAAGAGGAATTTCTAAGCTATTAGAAGCGGTACATTTTACCAGAAAGGCTAGCT 2100  
Db 681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaValHisLeuProLysArgLeuAla 700  
QY 2101 ATTATACACTGTCTCGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACACAGATG 2160  
Db 701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720  
QY 2161 GCTGACCGGTTGCCAGCAGGAGCCAGGGTGTAACTTCTGCTCTATAATAGAAATG 2220  
Db 721 AlaAspArgValAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProIleIleGluMet 740  
QY 2221 CCCAAAGCCCAAGAACCCAGACGACGTACACCTCAGAACACTGGCAAGAGATATAAAAG 2280  
Db 741 ProLysAlaProGluProArgArgGlnFyrThrLeuGluAspTrpGlnGluIleLysLys 760  
QY 2281 ATAGACCACTGTCTGAGACTCCGGAAGGACCTGTATATCCTCAGATGGGAAGGAATC 2340  
Db 761 IleAspGlnPheSerGluThrProGluGlyThrCysFyrThrSerAspGlyLysGluIle 780  
QY 2341 CTGCCCCACAAGAGGGTTAGATATGTCCAACAGATACATGCTTAACCCACCTAGGA 2400  
Db 781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
QY 2401 ACTAAACACCTCGACGAGTTGGTCAGAACATCCCTTATCATGTCTGAGGCTACCAGGA 2460  
Db 801 ThrIysHisLeuGlnGlnLeuValArgThrSerProFyrThrHisValLeuArgLeuProGly 820  
QY 2461 GTGGCTGACTCGGTGGTCAACACTTGTGTGCCCTGCCAGCTGGTTAATGCTAATCTCTCC 2520  
Db 821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
QY 2521 AGAATGCTCCAGGAAGAGACTAAGGGGAAGCCACCAGCGCTCATCGGAAGTGGAC 2580  
Db 841 ArgMetProProGlyLysArgLeuArgGlySerHisProGlyAlaHisIleTrpGluValAsp 860  
QY 2581 TTCACCTGAGTAAAGCCGGCTTAATACGGAAACAAATACCTATTGGTTTGTAGACACC 2640  
Db 861 PheThrGluValLysProAlaLysTyrGlyAsnLysFyrLeuLeuValPheValAspThr 880  
QY 2641 TTTTCAGGATGGGTAGAGGCTTATCTTACTAAGAAAGAGACTTCAACCGGTGGGTAAA 2700  
Db 881 PheSerGlyTrpValGluAlaFyrProThrLysLysGluThrSerThrValValAlaLys 900  
QY 2701 AAAATACTGGAAGAAATTTTCCAGATTTGGAAATACCTAAGGTAAATAGGGTCAGACAAT 2760

Db 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920  
QY 2761 GGTCCAGCTTTTGTTCGCCAGGTAAAGTCAGGACTGGCCCAAGATATTGGGATTTGATGG 2820  
Db 921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
QY 2821 AAACCTGCATTGTGTATACAGACCCCAAGCTCAGACACAGGTAGAGAGGATGAATAGAACC 2880  
Db 941 LysLeuHisCysAlaFyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
QY 2881 ATTTAAAGAGACCTTACTTAATTTGACCCGCGAGACTGCGGTTAATGATTGATAGCTCTC 2940  
Db 961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTrpIleAlaLeu 980  
QY 2941 CTGCCCTTTGTGCTTTTGTAGGTTAGGAACACCCCTGGACAGTTTGGCTCACCCTCAT 3000  
Db 981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProFyr 1000  
QY 3001 GAATTACTCTACGGGGGACCCCTTGTGTAGAAATTTGCTTGTACATAGTGTCTGAC 3060  
Db 1001 GluLeuLeuTyrGlyGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1020  
QY 3061 GTGCTGCTTCCACAGCTTTTGTCTTAGGCTCAAGGCACCTTGAGTGGGTGAGACAACGA 3120  
Db 1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
QY 3121 GCGTGGAGGCAACTCCGGGAGGCTACTCAGGAGGAGGACTTGCAGATCCCACATCGT 3180  
Db 1041 AlaTrpArgGlnLeuArgGluAlaFyrSerGlyGlyAspLeuGlnIleProHisArg 1060  
QY 3181 TTCCAAGTGGAGATTCAAGTCTAGTTCAGCCGCGCAGGAAACCTCGAGACTCGG 3240  
Db 1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
QY 3241 TGGAGGGCCCTTATCTCGTACTTTTGACCAACACCGCTGGAAGTGAAGGAAATC 3300  
Db 1081 TrpLysGlyProFyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1100  
QY 3301 TCCACCTGGATCCATGCATCCCGTTAAACCGCGCCACCTCCCGATTCGGGTGGAAA 3360  
Db 1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTrpLys 1120  
QY 3361 GCCGAAAGACTGAAATCCCTTAAGCTTTCGCTTCCATCGCTGGTTCCTTACTCTGTC 3420  
Db 1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProFyrSerVal 1140  
QY 3421 AATAACCTCTCAGAC 3435  
Db 1141 AsnAsnLeuSerAsp 1145  
RESULT 2  
AAB73286  
ID AAB73286 standard; protein; 1145 AA.  
XX AAB73286;  
XX  
XX 23-MAY-2001 (first entry)  
XX  
DE Retroviral protein #2 found in miniature swine.  
XX Retrovirus; graft transplantation; xenotransplantation; miniature swine.  
XX Unidentified.  
XX  
XX US6190861-B1.  
XX  
XX 20-FEB-2001.  
XX  
XX 13-DEC-1996; 96US-00766528.  
XX  
XX 14-DEC-1995; 95US-00572645.  
XX  
XX (GEO ) GEN HOSPITAL CORP.

```
XX Fishman JA;
XX
XX WPI; 2001-256211/26.
XX N-PSDB; AAF77727.
XX
XX Assessing risk of endogenous retroviruses in clinical practice and in
XX xenotransplantation, comprises using probe sequences derived from swine
XX or miniature swine retroviral genome.
XX
XX Disclosure; Fig 3; 127pp; English.
XX
XX The present invention relates to a method for screening a cell or tissue
XX for the presence or expression of a retrovirus (RV), comprising
XX contacting a target nucleic acid from the cell or tissue with a second
XX nucleic acid from the present invention (e.g. AAF7727 or a fragment
XX thereof). The method is useful for RV detection and to assess graft
XX transplation risk. Screening of animals allows the elimination of
XX donors with active replication of known viruses. Inactive proviruses can
XX be detected and inactivated, allowing identification and elimination of
XX potential human pathogens derived from swine in a manner not possible in
XX the outbred human organ donor population and is important to the
XX development of human xenotransplantation
XX
XX SQ Sequence 1145 AA;
XX
XX Alignment Scores:
XX Pred. No.: 0 Length: 1145
XX Score: 6038.00 Matches: 1145
XX Percent Similarity: 100.0% Conservative: 0
XX Best Local Similarity: 100.0% Mismatches: 0
XX Query Match: 97.7% Indels: 0
XX DB: 4 Gaps: 0
XX
XX US-10-723-552-3_COPY_2307_5741 (1-3435) x AAB73286 (1-1145)
XX
XX QY 1 ATGGGTGCCACAGCGGCAACACAGTATCCATGGACTACCCGAAGAACAGTTGACTTGGGA 60
XX DB 1 MetGlyAlaThrGlyGlnGlnGlnTrpTrpThrThrArgArgThrValAspLeuGly 20
XX
XX QY 61 GTGGGACGGGTAAACCACTCGTTTCTGTGTATACCTGAGTGCACGACCCTCTTAGGT 120
XX DB 21 ValGlyArgValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 40
XX
XX QY 121 AGAGACTATTGACCAAGATCGGAGCACAATTTCTTTGAACAGGGAACCAAGAGTG 180
XX DB 41 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyLysProGluVal 60
XX
XX QY 181 TCTGCAATAAACAACCTATCACTGTGTGTGACCTCCCAATTAGATGACGAATATCGACTA 240
XX DB 61 SerAlaAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyArgLeu 80
XX
XX QY 241 TACTCTCCCTAGTAAAGCCTGATCAAAATATACAATTCTGGTGGAAACAGTTTCCCAA 300
XX DB 81 TyrSerProLeuValLysProAspGlnAsnIleGlnPheTrpLeuGluGlnPheProGln 100
XX
XX QY 301 GCCTGGGAGAAAACCGCAGGAGATGGTTGGCAAGCAAGTTCCCCCAACAGTTATTCAA 360
XX DB 101 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 120
XX
XX QY 361 CTGAGGCCAGTGCACACACAGTGTCACTCAGACAGTACCCCTTGAGTAAAGAGACTCAA 420
XX DB 121 LeuLysAlaSerAlaThrProValSerValArgGlnTrpProLeuSerLysGluAlaGln 140
XX
XX QY 421 GAAGGAATTCGGCCGCATGTCCAAAGATTATCCAACAGGGCATCCTAGTTCTCTGCCAA 480
XX DB 141 GluGlyIleArgProHisValGlnArgLeuIleGlnGlnIleLeuValProValGln 160
XX
XX QY 481 TCTCTCCGGAATATCTCCCTGTACCGGTTAGAAAGCCTGGGACTAATGACTATCGACCA 540
XX DB 161 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTrpArgPro 180
XX
XX QY 541 GTACAGGACTTGAGAGAGTCAATAAACGGGTGCAGGATATACACCAACAGTCCCGAAC 600
XX
```

```
181 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 200
601 CTTTATAACCTCTTGTGTCTCTCCACCCCAACGGAGCTGTATACAGTATTGGACTTA 660
201 ProTyAsnLeuLeuCysAlaLeuProProGlnArgSerTrpTrpThrValLeuAspLeu 220
661 AAGGATGCCCTCTTCTTCCTGTGAGATTACACCCCACTAGCCAAACCACTTTTGGCTTCGAA 720
221 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 240
721 TGGAGAGATCCAGTACGGGAAGAACCGGCAGCTACCTGGAGCCCGACTGCCCAAGGG 780
241 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 260
781 TTCAGAAGCTCCCGACCATCTTTGCGAAGCCCTACACAGAGACCTGGCCCACTTCAGG 840
261 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 280
841 ATCCAAACACCTCAGGTGACCTCCTCCAGTACGTGATGACCTGTCTTCTGGCGGGAGCC 900
281 IleGlnHisProGlnValThrLeuLeuGlnTrpValAspAspLeuLeuLeuAlaGlyAla 300
901 ACCAAACAGACTCTCTTAGAAGGCACGAAGGCCTACTGCTGGAATTGTCTGACCTAGGC 960
301 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuSerAspLeuGly 320
961 TACAGACCTCTGCTAAGAGGCCAGATTTCGAGGAGAGGTTAAACATACTTTGGGTAC 1020
321 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTrpLeuGlyTyr 340
1021 AGTTTGGCGGACGGGACGATGGCTGACGAGGACGAGCAAGAAACAACTGTAGTCCAGATA 1080
341 SerLeuArgAspGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 360
1081 CCGGCCCAACCAACAGCCAAACAAATGAGAGAGTTTTCGGGACAGCTGGAATTTTGAGA 1140
361 ProAlaProThrThrAlaLysGlnMetArgGluPheLeuGlyThrAlaGlyPheCysArg 380
1141 CTGTGGATCCGGGGTTGGACCTTAGCAGCCCACTCTACCCGCTAAACCAAGAAAAA 1200
381 LeuTrpIleProGlyPheAlaThrLeuAlaProLeuTyrProLeuThrLysGluLys 400
1201 GGGGAATTCCTCGGGCTCTGAGCACAGAGCATTTGTATGCTATCAAAAAGCCCTG 1260
401 GlyGluPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 420
1261 CTGAGCGCACGTCTCTGGCCCTCCCTGACGTAACTAAACCTTTTACCTTTATGTGGAT 1320
421 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 440
1321 GAGGTAAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTTAGCAGCCATGGAGAGACCT 1380
441 GluArgLysGlyValAlaAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgPro 460
1381 GTCCCTACCTGTCAAAAGAGCTCGATCCTGTAGCCAGTGTGGCCCATATGCTCAAG 1440
461 ValAlaTyrLeuSerLysLysLysLeuAspProValAlaSerGlyTyrProIleCysLeuLys 480
1441 GCTATCGCAGCTGTGGCCCATCTGCTCAAGGACCTGACAAATTTGACTTTTGGGACAGAA 1500
481 AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 500
1501 ATAACCTTAATAGCCCCCATGATTCGATTCGAGAGACATCGTTCGGCAGCCCGCAGCGTGG 1560
501 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 520
1561 ATGACCAAGCCCGCAGTACCCCACTATCAAAAGCTGTCTTCTCACAGAGAGGCTCACGTTTC 1620
521 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 540
1621 GCTCCACAGCCGCTCTCAACCCCTGCCACTCTTCTGCTGAAGAGACTGATGAACCACTG 1680
```



Db 541 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 560  
QY 1681 ACTCATGATCCCACTCACTATTGATGAGGAGACTGGGGTCCGCAAGCACTTACAGAC 1740  
Db 561 ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValAaGlyAspLeuThrAsp 580  
QY 1741 ATACCGCTGACTGAGAAGTCTCAACCTGGTTCACTGACGGAAGACGATATGTGGTGAA 1800  
Db 581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 600  
QY 1801 GGTAAAGAGGATGGTGGGGCGGGTGGTGGACCGGACCCGACGATCTGGCCGACGAC 1860  
Db 601 GlyIysArgMetAlaGlyAlaAlaValAspGlyThrArgThrIleTrpAlaSerSer 620  
QY 1861 CTGCGGGAAGAACTTCAGCACAAAAGGCTGAGCTCATGGCCCTCAGCGCAAGCTTTGCGG 1920  
Db 621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
QY 1921 CTGCGGGAAGAAATCATAAACATTTATACGACAGCGATGATGCTTTGCGACTGCA 1980  
Db 641 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 660  
QY 1981 CACGTACATGGGCACTTATAACAAGGGGTGCTTACCTCAGCAGGGAGGGAATA 2040  
Db 661 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
QY 2041 AAGAACAAGAGGAAATTTCTAAGCTATTAGAAGCCGTACATTTTACCAAAAGGCTAGCT 2100  
Db 681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaValHisLeuProLysArgLeuAla 700  
QY 2101 ATTATACACTGCTCGACATCAGAAAGCTTAAGATCTCATATCCAGAGGAACCCAGATG 2160  
Db 701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720  
QY 2161 GCTGACCGGTTCGCAAGCAGGACCCAGGGTGTAAACCTTCTGCTTATATAGAAATG 2220  
Db 721 AlaAspArgValAlaLysGlnAlaIleGlnGlyValAsnLeuLeuProIleIleGluMet 740  
QY 2221 CCCAAAGCCCCAGAACCCAGACAGCAGTACACCTTAGAAGCTGGCAAGAGATAAAAAAG 2280  
Db 741 ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760  
QY 2281 ATAGACAGATTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGGAATC 2340  
Db 761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 780  
QY 2341 CTGCCCCAACAGAGAGGTTAGATATGTCCACAGATACATCGTTAAACCCACCTAGGA 2400  
Db 781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
QY 2401 ACTAAACACCTGCAGACATGGTTCAGAAACATCCCTTATCATGTTCTGAGGCTACAGGA 2460  
Db 801 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
QY 2461 GTGCGTCACTCGTGTGTCAAACATTTGTGTCCTGCCAGCTGGTTAATGCTAATCTCTCC 2520  
Db 821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
QY 2521 AGAATGCTCCAGGGAAGAGACTAAGGGGAAGCCACCGCGCTCACTGGGAAGTGGAC 2580  
Db 841 ArgMetProProGlyLysArgLeuArgGlySerHisProGlyAlaHisTrpGluValAsp 860  
QY 2581 TTCACCTGAGTAAGCGCGCTAAATACGGAACAAATACCTATTGTTTTTGTAGACACC 2640  
Db 861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
QY 2641 TTTTCAGGATGGGTAGAGGCTTATCTCTACTAAGAAAGAGACTTCAACCGTGGTGGTAAA 2700  
Db 881 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900  
QY 2701 AAAATACTGAAGAAATTTTCCAGATTTGGAAATACCTTAAGTAATAGGGTCAGACAAT 2760  
Db 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920

QY 2761 GGTCCAGCTTTGTTGCCAGGTAAAGTCAGGACTGCGCAAGATATTGGGATTGATTGG 2820  
Db 921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
QY 2821 AAACCTGATTGTGCATACAGACCCCAAGGCTCAGCAGGTAGAGAGATGAATAGAACC 2880  
Db 941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
QY 2881 ATTAAAGAGACCCTTACTAAATTGACCGCGGAGACTGCGGCTTAATGATTGATAGCTCTC 2940  
Db 961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTrpIleAlaLeu 980  
QY 2941 CTGCCCCCTTTGCTTTTGTAGGTTAGGAACACCCCTCGACAGTTTGGGTGACCCCTTAT 3000  
Db 981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
QY 3001 GAATTAATCTACCGGGGACCCCCCATTTGGTAGAAATTGCTCTGTACATAGTGTCTGAC 3060  
Db 1001 GluLeuLeuTyrGlyGlyProProProLeuValGluIleAlaSerValHisSerAlaAsp 1020  
QY 3061 GTGCTGCTTTCCAGCCTTTGTTCTTAGGCTCAAGCCTTGAAGTGGGTGAGACAAGGA 3120  
Db 1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
QY 3121 GCGTGGAGGCAACTCCGGGAGGCGCTACTCAGGAGGAGGAGACTTGCAGATCCCAATCGT 3180  
Db 1041 AlaTrpArgGlnLeuArgGluAlaTyrSerGlyGlyAspLeuGlnIleProHisArg 1060  
QY 3181 TTCCAAGTGGAGATTCAAGTCTAGTTAGACGCCACCGTCAGGAAAACCTCGAGACTCGG 3240  
Db 1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
QY 3241 TGGAAAGGCGCTTATCTCGTACTTTTGCACCAACCGCTGTGAAGTCGAGGAATC 3300  
Db 1081 TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1100  
QY 3301 TCCACCTGGATCCATGCATCCACGTTAAACCCGCGCACCTCCCGATTCCGGGTGGAAA 3360  
Db 1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProProAspSerGlyTrpLys 1120  
QY 3361 GCCCAAGAAGACTGAAATCCCTTAAAGCTTCGCTCCATCGCTGGTTCCTTACTCTGTC 3420  
Db 1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1140  
QY 3421 AATAACTCTCAGAC 3435  
Db 1141 AsnAsnLeuSerAsp 1145  
RESULT 3  
AD573447  
ID AD573447 standard; protein; 1145 AA.  
XX  
AC AD573447;  
XX  
DT 16-DEC-2004 (first entry)  
XX  
DE Swine retroviral pol protein.  
XX  
KW Swine retroviral protein; immunosuppressive; gene therapy; pol protein.  
XX  
OS Porcine endogenous retrovirus.  
XX  
PN US2004185435-A1.  
XX  
PD 23-SEP-2004.  
XX  
PF 26-NOV-2003; 2003US-00723552.  
XX  
PR 14-DEC-1995; 95US-00572645.  
PR 13-DEC-1996; 96US-00766528.  
PR 14-SEP-2000; 2000US-00661858.  
XX



Db 541 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 560  
Qy 1681 ACTCATGATTCCTCACTATTGATTGAGGAGACTGGGGTCCGCAAGAGACTTACAGAC 1740  
Db 561 ThrHisAspCysHisGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp 580  
Qy 1741 ATACCGCTGACTGAGAAGTCTAACTGCTTCTACTGACGAAGCAGCTATGCTGGTGA 1800  
Db 581 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 600  
Qy 1801 GGTAAAGAGGATGGCTGGGGCGGCTGGTGACCGGACCCGACAGATCTGGGGCCAGCAGC 1860  
Db 601 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 620  
Qy 1861 CTGCGCGAAGAACTTCAGCACAAAGCTGAGCTCATGGCCCTCACGCAAGCTTTCGG 1920  
Db 621 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 640  
Qy 1921 CTGCGCGAAGAAATCCATAAACATTTATACGACAGCAGCTATGCTTTCGCACTGCA 1980  
Db 641 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 660  
Qy 1981 CACTTACATGGGGCATCTATAAACAAGGGGTTGCTTACTCTCAGCAGGAGGGAATA 2040  
Db 661 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 680  
Qy 2041 AAGAACAAAGAGAAATCTTAAGCTATTAGAAGCGGTACATTACCAAAAAGCTAGCT 2100  
Db 681 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaValHisLeuProLysArgLeuAla 700  
Qy 2101 ATTATACATCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGAAACAGATG 2160  
Db 701 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 720  
Qy 2161 GCTGACCGGTTGCAAGCGGACCCAGGGTCTTAACCTTCTGCTTATAATAAGAAATG 2220  
Db 721 AlaAspArgValAlaLysGlnAlaAlaGlnGlyValAsnLeuLeuProIleIleGluMet 740  
Qy 2221 CCCAAAGCCCCAGAACCCAGACGACAGTACACCCCTAGAAAGACTGGCAAGAGATAAAAAG 2280  
Db 741 ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys 760  
Qy 2281 ATAGACAGTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAAATC 2340  
Db 761 IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerAspGlyLysGluIle 780  
Qy 2341 CTGCCCCACAAAGAGGTTAGAAATGTCCAACAGATACATCGCTAAACCCACCTAGGA 2400  
Db 781 LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly 800  
Qy 2401 ACTAAACACCTGCAGCAGTTGGTCAGAACATCCCTTATCATGTTCTGAGGCTACCAAGGA 2460  
Db 801 ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly 820  
Qy 2461 GTGCTGACTCGGTGGTCAACATTTGTTGTCCTGCCAGCTGGTGAATGTAATCTCTTCC 2520  
Db 821 ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer 840  
Qy 2521 AGAATGCTCCAGGAGAGACTAAGGGAGCCAGCCAGCGCTCACTGGGAAGTGAC 2580  
Db 841 ArgMetProProGlyLysArgLeuArgGlySerHisProGlyAlaHisIleTrpGluValAsp 860  
Qy 2581 TTCCTCAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGGTTTTTGTAGACACC 2640  
Db 861 PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr 880  
Qy 2641 TTTTCAGATGGGTAGAGCTTATCTCTACTAAGAAAGAGACTTCAACCGGTGGCTAAA 2700  
Db 881 PheSerGlyTrpValGluAlaTyrProThrLysLysGluThrSerThrValValAlaLys 900  
Qy 2701 AAAATACTGGAAGAAATTTTCCAGATTTGGATACCTAAGCTAAGGTATAGGGTCAGCAAT 2760  
Db 901 LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn 920

Qy 2761 GGTCCAGCTTTTGTGCCAGTAAAGTACAGGACTGGCCAAGATATTGGGATTTGATTGG 2820  
Db 921 GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp 940  
Qy 2821 AAACCTGATTTGTGATACAGACCCCAAGACTCAGACAGGTAGAGAGATGAATAGAAC 2880  
Db 941 LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr 960  
Qy 2881 ATTAAAGAGACCTTTACTAAATTGACCGGAGACTGGCGTTAATGATTGATAGCTCTC 2940  
Db 961 IleLysGluThrLeuThrLysLeuThrAlaGluThrGlyValAsnAspTrpIleAlaLeu 980  
Qy 2941 CTGCCCCCTTTGTGCTTTTGTAGGTTAGAAACACCCCTCGACAGTTTGGGCTGACCCCTAT 3000  
Db 981 LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr 1000  
Qy 3001 GAATTACTCTACGGGGACCCCGCCCATTTGGTAGAATAATGCTTCTGTACATAGTCTGAC 3060  
Db 1001 GluLeuLeuTyrGlyGlyProProLeuValGluIleAlaSerValHisSerAlaAsp 1020  
Qy 3061 GTGCTGCTTTCCAGACCTTTGTTCTTAGGCTCAAGGACCTTGAGTGGGTGAGACACGA 3120  
Db 1021 ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluTrpValArgGlnArg 1040  
Qy 3121 GCGTGGAGGCAACTCCGGGAGGCTCTCTCAGGAGGAGGAGACTTGCAGATCCACATCGT 3180  
Db 1041 AlaTrpArgGlnLeuArgGluAlaTyrSerGlyGlyAspLeuGlnIleProHisArg 1060  
Qy 3181 TTCCAAGTGGAGATTCTAGTCTAGTACGCCACCGTCAGAGAAACCTCGAGACTCGG 3240  
Db 1061 PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg 1080  
Qy 3241 TGGAAAGGCGCTTATCTCTGTTTACCCACACCGCTGTGAAGTCAAGGAATC 3300  
Db 1081 TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle 1100  
Qy 3301 TCCACCTGGATTCATGCCACCTTAAACCGCGCCACCTCCCGATTTCGGGGTGGAAA 3360  
Db 1101 SerThrTrpIleHisAlaSerHisValLysProAlaProProAspSerGlyTrpLys 1120  
Qy 3361 GCGGAAAGACTGAAATCCCTTAAGCTTCGCTCCATCGGTGGTTCCTTACTCTGTC 3420  
Db 1121 AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgValValProTyrSerVal 1140  
Qy 3421 AATAACCTCTCAGAC 3435  
Db 1141 AsnAsnLeuSerAsp 1145

## RESULT 4

AAW39272  
ID AAW39272 standard; protein; 1194 AA.

XX AAW39272;

XX 27-AUG-2003 (revised)

XX 19-MAY-1998 (first entry)

XX Porcine retrovirus POL protein.

XX Porcine retrovirus; PoEV; POL protein; ENV protein; GAG protein; vaccine;  
diagnosis; xenotransplantation; prophylactic; therapeutic.

XX Pig endogenous retrovirus.

XX WO9740167-A1.

XX 30-OCT-1997.

XX 18-APR-1997; 97WO-GB001087.

XX 19-APR-1996; 96GB-00008164.

PR 10-FEB-1997; 97GB-00002668.

XX (QONE-) Q-ONE BIOTECH LTD.  
PA (IMUT-) IMUTRAN LTD.  
XX Galbraith DN, Haworth C, Lees GM, Smith KT;  
XX WPI; 1997-535851/49.  
DR N-PSDB; AAV03700.  
XX Polynucleotide encoding porcine retrovirus expression product - useful to  
PT develop products for use in vaccines, diagnosis and xeno-transplantation.  
XX Claim 6; Fig 3; 69pp; English.  
XX This sequence represents the porcine retrovirus (PoEV) polymerase (POL)  
CC protein. This protein and other porcine retroviral proteins e.g. the  
CC vision core (CAG) and envelope (ENV) proteins can be used to develop  
CC viral vaccines, antisense nucleic acids, ribozymes and other antiviral  
CC agents. They can also be used in xeno-transplantation technology and as  
CC diagnostic tools. (Updated on 27-AUG-2003 to correct OS field.)  
XX  
SQ Sequence 1194 AA;  
Alignment Scores:  
Pred No.: 0 Length: 1194  
Score: 5851.50 Matches: 1107  
Percent Similarity: 98.4% Conservative: 17  
Best Local Similarity: 96.9% Mismatches: 17  
Query Match: 94.6% Indels: 1  
DB: 2 Gaps: 1  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x AAW39272 (1-1194)  
QY 1 ATGGGTGCCAGGCAACACAGTATCCATGAGTACCCGAGAACAGTTGACTTGGGA 60  
DB 50 MetGlyAlaThrGlyGlnArgGlnTyrProTrpThrThrArgArgThrValAspLeuGly 69  
QY 61 GTGGGAGGGTAACCCACTCGTTCTGTGTATCTAGTGTGCGGACCCAGCCCTCTAGGT 120  
DB 70 ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly 89  
QY 121 AGAGACTTATTGACCAAGATGGGAGCACAAATTTCTTTTGAACAGGGAACACAGAGTG 180  
DB 90 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal 109  
QY 181 TCTGCAATAACAAACCTATCACTGTGTGTGACCTTCCCAATTAGATGAGCAATATCGACTA 240  
DB 110 SerValAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyrArgLeu 129  
QY 241 TACTCTCCCTAGTAAAGCCTGATCAAAATATACAATTTCTGTTGGACAGTTTCCCCAA 300  
DB 130 TyrSerProGlnValLysProAspGlnAspIleGlnSerTrpLeuGlnGlnPheProGln 149  
QY 301 CCCTGGGAGAAACCGCAGGATGGTTTGGCAAGCAAGTTCCCCCAAGTATTTCAA 360  
DB 150 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProGlnValIleGln 169  
QY 361 CTGAAGGCCAGTGCACACAGTGTCACTCAGACAGTACCCCTTGAGTAAAGAGCTCAA 420  
DB 170 LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaArg 189  
QY 421 GAAGGAATTCGGCGCATGTCCAAAGATTATCCACAGGGCATCTAGTTCTGTGCCAA 480  
DB 190 GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 209  
QY 481 TCTCCCTGGAATACTCCCTGCTACCGGTAGAAAGCCTGGGACTTAATGACTATCGACCA 540  
DB 210 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 229  
QY 541 GTACAGGACTTGAGAGGTCATTAACGGGTGACAGGATATACCCCAACAGTCCCGAAC 600  
DB 230 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 249

QY 601 CTTTATAACCTCTGTGTCTCTCCACCCCAACGAGCTGGTATACAGTATTGGACTTA 660  
DB 250 ProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeu 269  
QY 661 AAGGATCCCTTCTTCTGCTGAGATTACCCCCACTAGCCAAACCACTTTTTCCTTCGAA 720  
DB 270 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 289  
QY 721 TGGAGAGATCCAGTACGGGAGAACCGGCAGCTCACCTGGACCCGACTGCCCCACAGGG 780  
DB 290 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 309  
QY 781 TTCAGAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGG 840  
DB 310 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 329  
QY 841 ATCCAACACCTCTAGGTGACCCCTCTCAGTACGTGATGACCTGCTTCTGGCGGAGCC 900  
DB 330 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuAlaGlyAla 349  
QY 901 ACCAAACAGGACTCTTAGAAGCAGGACGACACTACTGCTGGAAATTTGCTGACTAGGC 960  
DB 350 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuGluLeuSerAspLeuGly 369  
QY 961 TACAGACCTCTGCTAAGAAGGCCAGATTTCAGGAGAGAGGTAAACATACTTCGGGTAC 1020  
DB 370 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 389  
QY 1021 AGTTTGGCGGACGGGACGATGGCTCACCGAGGACGCGAAGAAACCTGTAGTCCAGATA 1080  
DB 390 SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 409  
QY 1081 CCGGCCCCAACCCACAGCAACAAATGAGAGAGTTCCTGGGACAGCTGGATTTTGCAGA 1140  
DB 410 ProAlaProThrThrAlaLysGlnValArgGlnPheLeuGlyThrAlaGlyPheCysArg 429  
QY 1141 CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCACTTACCCGCTAAACCAAGAAAAA 1200  
DB 430 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrLysGluLys 449  
QY 1201 GGGGAATTTCTCTGGGCTCTGAGCACAGAGCATTTGATGCTATCAAAAAGCCCTG 1260  
DB 450 GlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 469  
QY 1261 CTGAGCCGACCTGCTCTGGCCCTCCCTGAGCTAAACCCCTTACCTTTATGTGGAT 1320  
DB 470 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 489  
QY 1321 GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCCAAACCCCTAGGACCATGAGAGACCT 1380  
DB 490 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 509  
QY 1381 GTCCCTACTGTCTCAAGAGAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCTGAAG 1440  
DB 510 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTyrProValCysLeuLys 529  
QY 1441 GCTATCGAGCTGTGGCCCATCTGCTCAAGACCCCTGCACAAATTCACCTTTCGGACAGAT 1500  
DB 530 AlaIleAlaAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 549  
QY 1501 ATAACTGTAAATAGCCCCCATGATGGAGAAACATCGTTCCGACAGCCCCCAGACCGATGG 1560  
DB 550 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 569  
QY 1561 ATGACCAACGCCCATGACCCACTATCAAGCTGCTTCTTCACAGAGAGGGTCAACGTTTC 1620  
DB 570 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 589  
QY 1621 GCTCCACAGCCGCTCTCAACCCCTGCCACTCTCTGCTGAGAGACTGTATGACCACTG 1680  
DB 590 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuLeuProGluGluThrAspGluProVal 609  
QY 1681 ACTCATGATTGCCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTTACAGAC 1740

Db	610	ThrHisAspCysHisIleGlnLeuLeuIleGluGluThrGlyValArgLysAspLeuThrAsp	629
QY	1741	ATACCGCTACTGGAGAAAGTGTAACTCGTGGTTCACTGACGGAAGCAGCTATGTGTGGTCAA	1800
Db	630	IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu	649
QY	1801	GGTAAGAGGATGGCTGGGGCGCGGTGTGTGACGGGACCCGACGATCTGGGCCACGACG	1860
Db	650	GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer	669
QY	1861	CTGGCGGAAGGAACCTTCAGCACAAAGCTGAGCTCATGGCCCTCACGCAAGCTTTGGCG	1920
Db	670	LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg	689
QY	1921	CTGGCCGGAAGGAATCCATAAAACATTTATACGGACAGCGATGCTCTTTGCCACTGCA	1980
Db	690	LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla	709
QY	1981	CACGTACATGGGGCCCATCTATAACAAGGGGGTTGCTTACCTCAGCAGGGAGGGAATA	2040
Db	710	HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle	729
QY	2041	AAGACAAAGGGAATCTTAAGCCTATTAGACCGGTACATTTACAAAAAGCGTAGCT	2100
Db	730	LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla	749
QY	2101	ATTATACACTGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAACACGATG	2160
Db	750	IleIleHisCysProGlyHisIleGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet	769
QY	2161	GCTGACCGGGTTGCCAAGCAGCGCCAGCGGTGTAACTTCTGCCCTATATAGAATG	2220
Db	770	AlaAspArgValAlaLysGlnAlaAlaGlnAlaValAsnLeuLeuProIleIleGluThr	789
QY	2221	CCCCAAGCCCCAGAACCCAGACGACGTACACCCCTAGAGACTGGCAAGAGATAAAAAG	2280
Db	790	ProLysAlaProGluProArgArgGlnTyrThrLeuGluAspTrpGlnGluIleLysLys	809
QY	2281	ATAGACAGATTCTCTGAGACTCCCGAAGGACCTGCTATACCTCAGATGGGAAGGAATC	2340
Db	810	IleAspGlnPheSerGluThrProGluGlyThrCysTyrThrSerTyrGlyLysGluIle	829
QY	2341	CTGCCCCCAAGAAAGGGTTAGATAATGTCCAACAGATACATCGCTAAACCCACTAGGA	2400
Db	830	LeuProHisLysGluGlyLeuGluTyrValGlnGlnIleHisArgLeuThrHisLeuGly	849
QY	2401	ACTAAACACCTGCAGCAGTTGGTCAGAACATCCCTTATCATGTTCTGAGGCTACCAAGA	2460
Db	850	ThrLysHisLeuGlnGlnLeuValArgThrSerProTyrHisValLeuArgLeuProGly	869
QY	2461	GTGGCTCACTCGGTGGTCAAAACATTGTGTGCCCTCCAGCTGGTGTATGCTATCTCTCC	2520
Db	870	ValAlaAspSerValValLysHisCysValProCysGlnLeuValAsnAlaAsnProSer	889
QY	2521	AGAACTGCTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGCGCTCACTGGGAAGTGGAC	2580
Db	890	ArgIleProProGlyLysArgLeuArgLysSerHisProGlyAlaHisTrpGluValAsp	909
QY	2581	TTCACTAGGTAAAGCCGGCTAAATACGGAAACAAATACCTATTGTTTGTGTAGACACC	2640
Db	910	PheThrGluValLysProAlaLysTyrGlyAsnLysTyrLeuLeuValPheValAspThr	929
QY	2641	TTTTTCAGATGGGTAGAGGCTTATCTCTACTAAGAAAGAGACTTCAACCGTGGTGGCTAAA	2700
Db	930	PheSerGlyTrpValGluAlaIleTyrProThrLysLysGluThrSerThrValValAlaLys	949
QY	2701	AAAATACTGGAAGAAAATTTTTTCCAAGATTTGGGAATACCTAAGGTAATAGGTCAGACAAT	2760
Db	950	LysIleLeuGluGluIlePheProArgPheGlyIleProLysValIleGlySerAspAsn	969
QY	2761	GGTCCAGCTTTGTGTGCCCAGGTAAAGTCAGGGAAGTGGCCCAAGATATTGGGGAATTGATGG	2820

Db	970	GlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaLysIleLeuGlyIleAspTrp	989
QY	2821	AAACTGCATTGTGCATACAGACACCCAAAGCTCAGACAGGTAGAGAGGATGAATAGAAC	2880
Db	990	LysLeuHisCysAlaIleArgProGlnSerSerGlyGlnValGluArgMetAsnArgThr	1009
QY	2881	ATTAAGAGACCCCTTACTTAATTCACCCGGAGACTGGCGTTAATGATTTGGATGCTCTC	2940
Db	1010	IleLysGluThrLeuThrLysLeuThrThrGluThrGlyIleAsnAspTTPMetAlaLeu	1029
QY	2941	CTGCGCTTTGTGCTTTTGTAGGTTAGAAACACCCCTGACAGTTTGGGCTGACCCCTAT	3000
Db	1030	LeuProPheValLeuPheArgValArgAsnThrProGlyGlnPheGlyLeuThrProTyr	1049
QY	3001	GAATTACTCTACGGGGACCCCGGAGTGGTAGAAATTCCTCTGTACATATAGTCTGAC	3060
Db	1050	GluLeuLeuTyrGlyGlyProProProLeuAlaGluIleAlaPheAlaHisSerAlaAsp	1069
QY	3061	GTGCTGCTTTCCAGCCCTTTTCTTAGGCTCAAGGCACCTTGAGTGGGTGAGACACGA	3120
Db	1070	ValLeuLeuSerGlnProLeuPheSerArgLeuLysAlaLeuGluThrValArgGlnArg	1089
QY	3121	CGGTGGAGGCACTCCGGAGGCTCTACTCAGGAGGAGACTTGCAGATCCACATCGT	3180
Db	1090	AlaTrpLysGlnLeuArgGluAlaTyrSer---GlyGlyAspLeuGlnValProHisArg	1108
QY	3181	TTCCAAGTGGGAGATTCAAGTCTACGTTAGACGCCCGTCAGGAAACCTCGAGACTCGG	3240
Db	1109	PheGlnValGlyAspSerValTyrValArgArgHisArgAlaGlyAsnLeuGluThrArg	1128
QY	3241	TGGAAAGGCCCTTATCTCGTACTTTTGACCACACCAACCGCTGTGAAGATCGAAGGATC	3300
Db	1129	TrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysValGluGlyIle	1148
QY	3301	TCCACCTGGATCCATGTCATCCAGCTTAACCGCGCCACCTCCGATTCGGGGTGGAA	3360
Db	1149	ProThrTrpIleHisAlaSerHisValLysProAlaProProProAspSerGlyTrpArg	1168
QY	3361	GCCGAAAGACTGAAATCCCTTAAGCTTCGCTCCATCGCTGGTTCCTTACTCTGTC	3420
Db	1169	AlaGluLysThrGluAsnProLeuLysLeuArgLeuHisArgLeuValProTyrSerAsn	1188
QY	3421	AATAAC 3426	
Db	1189	AsnAsn 1190	
RESULT 5			
AAW81572	ID	AAW81572 standard; protein; 1203 AA.	
XX	AC	AAW81572;	
XX	DT	01-MAR-1999 (first entry)	
XX	DE	Mus dunni endogenous virus Pol protein.	
XX	DE	MDEV; retrovirus; packaging cell line; gene transfer; gene therapy;	
KW	KW	vector; polymerase protein; Pol protein.	
XX	OS	Mus dunni endogenous virus.	
XX	PN	WO9850538-A1.	
XX	PD	12-NOV-1998.	
XX	PF	08-MAY-1998; 98WO-US009452.	
XX	PR	09-MAY-1997; 97US-0046140P.	
XX	PR	08-MAY-1998; 98US-00075272.	
XX	FA	(HUTC-) HUTCHINSON CANCER RES CENT FRED.	
XX	PI	Miller AD, Wolgamot G, Bonham LJ;	

XX WPI; 1999-034718/03.  
 DR N-PSDB; AAV69750.  
 XX  
 PT New retroviral packaging cells - containing Mus dunni endogenous virus  
 PT sequences to target cells, retrovirus gag and pol genes and a  
 PT heterologous gene of interest.  
 XX  
 PS Disclosure; Page 68-71; 85pp; English.  
 XX  
 CC This is the amino acid sequence of the polymerase (Pol) protein of Mus  
 CC dunni endogenous virus (MDEV), as deduced from the MDEV nucleotide  
 CC sequence (see AAV69750). A cultured packaging cell is claimed which  
 CC produces a replication-defective retroviral vector (RDRV) particle, where  
 CC the packaging cell is a vertebrate cell capable of expressing and  
 CC assembling retroviral proteins, comprising: (a) a first vector encoding a  
 CC retroviral envelope protein having amino acid residues MDEV that direct  
 CC binding of the retroviral particle to MDEV retroviral receptors on a  
 CC target cell; and (b) a second vector encoding retrovirus Gag and Pol  
 CC proteins, where upon expression of the vectors in the packaging cell in  
 CC the presence of a vector having a sequence of a heterologous gene of  
 CC interest, a replication-defective retroviral particle is produced that  
 CC binds to MDEV receptors of target cells. Also claimed are: (1) a cultured  
 CC packaging cell for producing a RDRV particle; (2) methods for producing a  
 CC RDRV particle comprising a heterologous gene of interest; (3) cultured  
 CC packaging cell line PD223; and (4) a RDRV produced by a method as in (2).  
 CC The MDEV receptor is present on a variety of cells rendering MDEV  
 CC pseudotype packaging cells useful in methods of mammalian and  
 CC particularly human gene transfer for gene therapy. The MDEV packaging  
 CC cells are a stable and reproducible source of retroviral particles.  
 CC Clones may be isolated from these populations that produce high titre  
 CC virus. The packaging cell lines may be selected and cloned for other  
 CC desirable properties, such as stability of in vivo growth, lack of  
 CC production of helper virus, lack of reinfection by viral particles  
 CC packaged in the cell, stability from genetic rearrangement and  
 CC recombinational events, resistance to complement lysis, and improved  
 CC ability to infect cells from higher mammals  
 XX  
 SQ Sequence 1203 AA;

## Alignment Scores:

Pred. No.:	0	Length:	1203
Score:	4146.50	Matches:	786
Percent Similarity:	79.6%	Conservative:	125
Best Local Similarity:	68.6%	Mismatches:	213
Query Match:	67.1%	Indels:	21
DB:	2	Gaps:	7

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x AAW81572 (1-1203)

Qy	1	ATGGGTGCCACAGGGCAACACAGTATCCATGGACTACCCGAGAACACAGTTGACTTGGGA	60
Db	50	ValGlyAlaThrGlySerLysLeuTyrProTrpThrThrLysArgAlaLeuGlnIleasp	69
Qy	61	GTGGCAGGGTAACCCACTCGTTCTGTGTATCTGAGTGCCAGCACCCCTCTTAGGT	120
Db	70	LysAsnMetValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly	89
Qy	121	AGAGACTTATTGACCAAGATGGGACACAAATTTCTTTTGAACAGGGGAACACAGAAAGTG	180
Db	90	ArgAspLeuLeuThrLysLeuLysAlaGlnValGlnPheThrSerGluGlyProGlnVal	109
Qy	181	TCTGCAATAACAACCTATCACTGTGTGGTGGACCCCTCAATTAGATGACGATATCGACTA	240
Db	110	SerTrpGlyLysAlaProLeuAlaCysLeuValLeuSerThrGluGluGluTyrArgLeu	129
Qy	241	TACTCTCCCTAGTAAGCCGTGACAAATATCAATTCGTGTTGGAACAGTTCCTCCCAA	300
Db	130	HisGluGluGlnProLysGlyAlaAlaProLeuAsp---TrpValThrAlaPheProAsn	148
Qy	301	GCCTGGGAGAAACCGCAGGGATGGTTTGGCAAGCAAGTTCCTCCCAAGTATTATCAA	360
Db	149	ValTrpAlaGluGlnAlaGlyMetGlyLeuAlaLysGlnValProProValValValGlu	168

Qy	361	CTGAAGGCCAGTCCACACAGTGTCTCAGTCAGACAGTACCCCTTGAGTAAAGAAGCTCAA	420
Db	169	LeuLysAlaAspAlaThrProIleSerValArgGlnTyrProMetSerLysGluAlaLys	188
Qy	421	GAAGAAATTCGGCGCATGTCCAAAGATTAAATCAACAGAGGCATCTAGTTCCTGTCCAA	480
Db	189	GluGlyIleArgProHisIleArgArgLeuLeuAspGlnGlyIleLeuValAlaCysGln	208
Qy	481	TCTCCCTGGAAATACCTCCCTGCTACCGGTTAGAAAGCTGGGACTAATGACTATCGACCA	540
Db	209	SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro	228
Qy	541	GTACAGCACTTGAGAGAGGTCAATAACGGGTGAGGATATACACCAACAGTCCCCGAAC	600
Db	229	ValGlnAspLeuArgGluValAsnLysArgValLeuAspIleHisProThrValProAsn	248
Qy	601	CCTTATACCTCTTGTGTGCTCTCCCAACCCGAGCTGGTATACAGTATTGGACTTA	660
Db	249	ProTyrAsnLeuLeuSerSerLeuProGluArgThrTrpTyrThrValLeuAspLeu	268
Qy	661	AAGGATCCCTTCTCTCCCTGAGATTACACCCACTAGCCACACCTTTTGCCTTCGAA	720
Db	269	LysAspAlaPhePheCysLeuArgLeuHisProLysSerGlnLeuLeuPheAlaPheGlu	288
Qy	721	TGGAGAGATCCAGTACGGGAAGAACCGGCGAGCTCACCTGGACCCGACTGCCCCCAAGGG	780
Db	289	TrpArgAspProGluGlyGlyGlnThrGlyGlnLeuThrTrpThrArgLeuProGlnGly	308
Qy	781	TTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCCACTTCAGG	840
Db	309	PhelysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaProPheArg	328
Qy	841	ATCCAAACCTCCTCAGGTGACCTCTCCTCAGTACGTGATGACCTGCTTCTGGCGGAGCC	900
Db	329	AlaGlnAsnProGlnLeuThrLeuLeuGlnTyrValAspAspLeuLeuIleAlaAla	348
Qy	901	ACCAAAACAGGACTGCTTAGAAGCACCAAGGCACACTGCTGGAATTTGCTGACCTAGGC	960
Db	349	SerLysGluLeuCysGlnGlnGlyThrGluArgLeuLeuThrGluLeuGlyAsnLeuGly	368
Qy	961	TACAGAGCCTCTGCTAAGAAGGCCAGATTTCGACGAGAGAGAGTAACATCTTGGGTAC	1020
Db	369	TyrArgValSerAlaLysLysAlaGlnIleCysGlnThrGluValIleTyrLeuGlyTyr	388
Qy	1021	AGTTTGGCGACGGCGAGCATGCTGACGAGGACCGAAGAAACCTGAGTCCAGATA	1080
Db	389	ThrLeuArgGlyGlyLysArgTrpLeuThrGluAlaArgLysLysThrValMetMetIle	408
Qy	1081	CCGCCCCCAACACAGCCAAACAAATGAGAGAGTTCCTTTGGGACAGCTGGATTTTCGAGA	1140
Db	409	ProProProThrThrProArgGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg	428
Qy	1141	CTGTGGATCCCGGGTTTGGACCTTAGACGCCCTCTACCCGCTAACCAAGAAAA	1200
Db	429	LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrArgGluGly	448
Qy	1201	GGGGAATTCCTGGCTCTGACGACGAGGACATTTGATGCTATCAAAAAGGCCCTG	1260
Db	449	IleProPheGluTrpLysGluGluHisGlnArgAlaPheGluAlaIleLysSerSerLeu	468
Qy	1261	CTGAGGCACCTGCTCTGGCCCTCCCTGACGTAATAACCTTTTACCTTTATGTGGAT	1320
Db	469	MetThrAlaProAlaLeuAlaLeuProAspLeuThrLysSerPheValLeuTyrValAsp	488
Qy	1321	GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTTAGCACCATGAGAGACCT	1380
Db	489	GluArgAlaGlyIleAlaArgGlyValLeuThrGlnAlaLeuGlyProTrpLysArgPro	508
Qy	1381	GTCCCTACCTGTCAAAGAGCTGATCCTGTACCCAGTGGTTGGCCCATATGCTGAG	1440
Db	509	ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProThrCysLeuLys	528





XX RFB retrovirus; gag; pol; env; osteogenesis; osteoinductive protein;  
KW bone development; osteoporosis; gene therapy; polymerase;  
KW reverse transcriptase.  
XX Murine leukemia virus.  
XX DE4411718-A1.  
XX 12-OCT-1995.  
XX 05-APR-1994; 94DE-04411718.  
XX 05-APR-1994; 94DE-04411718.  
XX (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.  
XX Schmidt J, Gimbel W, Strauss P, Erfle V, Pedersen FS, Pedersen L;  
PI Oestergaard M;  
PI WPI; 1995-352078/46.  
DR N-ESDB; AAQ94266.  
XX RFB-14 retrovirus genome - and prodn. of osteo-inductive proteins.  
XX Claim 14; Fig 1; 46pp; German.  
XX The full-length proviral genomic sequence of retrovirus RFB-14 has been  
CC determined. The virus codes for an osteoinductive protein, although the  
CC precise location of the coding region has not yet been identified. The  
CC virus may be useful in gene therapy of bone growth disorders such as  
CC osteoporosis. The present sequence is that of the viral pol gene product.  
CC (Updated on 27-AUG-2003 to correct OS field.)  
XX SQ Sequence 1196 AA;

Alignment Scores:  
Pred. No.: 0 Length: 1196  
Score: 3818.00 Matches: 721  
Percent Similarity: 76.9% Conservative: 162  
Best Local Similarity: 62.8% Mismatches: 251  
Query Match: 61.7% Indels: 14  
DB: 2 Gaps: 9

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x AAR75189 (1-1196)

Qy 4 GGTGCCAGGGCAACACAGTATCCATGGACTACCCGAGAACACAGTTTGAGTGGAGTG 63  
Db GlyAlaThrGlyGlyLysArgTyrArgTrpThrThrAspArgLysValHisLeuAlaThr 70  
Qy 64 GGACGGGTAAACCACTCGTTCTGGTCACTAGTGGTCCAGCACCCTCTTAGGTAGA 123  
Db GlyLysValThrHisSerPheLeuHisValProAspCysProTyrProLeuLeuGlyArg 90  
Qy 124 GACTATTGACAGATGGGAGCAAAATTTCTTTGAA---CAAGGAAACCAAGAGTG 180  
Db AspLeuLeuThrLysLeuLysValGlnIleHisPheGluGlySerGlyAlaGlnValVal 110  
Qy 181 TCTGMAATAACAACCTATCACTGTGTGTGACCTCCCAATTAGATGACCAATATCGACTA 240  
Db GlyProLysGlyGlnProLeuGlnValLeuThrLeuAsnLeuGluAspGluTyrArgLeu 130  
Qy 241 TACTCTCCCTAGTAAAGCCTGATCAAAATATA---CAATTCTGGTTGGAAACAGTTTCCC 297  
Db TyrGluThrSerAlaGluProGluAlaSerProGlySerThrTrpLeuSerAspPhePro 150  
Qy 298 CAACCTGGGAGAAACCCAGGGATGGTTTGGCAAGCAAGTCCCCCAAGTATT 357  
Db GlnAlaTrpAlaGluThrGlyGlyMetGlyLeuAlaValArgGlnAlaProLeuIleIle 170  
Qy 358 CAATGAAGCCAGTGCACACCACTGTCAGTTCAGTACAGACAGTACCCCTTGAGTAAAGAGCT 417  
Db ProLeuLysAlaThrSerThrProValSerIleLysGlnTyrProMetSerGlnGluAla 190

Qy 418 CAAGAAGGAATTCGGCCGCATGTCCTCAAGATTATCCACAGGGCATCTCTAGTTCTCTGTC 477  
Db ArgLeuGlyIleLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys 210  
Qy 478 CAATCTCCCTGGAATACCTCCCTGCTACCGGTTAGAAAGCTGGGACTAATGACTATCGA 537  
Db GlnSerProTrpAsnThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg 230  
Qy 538 CCAGTACAGGACTTGAAGAGGTCATAAACGGGTGCAGGATATACACCCACAGTCCCG 597  
Db ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro 250  
Qy 598 AACCTTTATAACCTCTTGTGTCTCCCAACCCCAACGGAGCTGTATACAGTATTGGAC 657  
Db AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisArgTrpTyrThrValLeuAsp 270  
Qy 658 TTAAGGATGCTCTTCTCTGCTGAGATTACACCCCTAGCCCAACCACTTTTTCCTTC 717  
Db LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPhe 290  
Qy 718 GAATGAGAGATCCAGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTCCCA 777  
Db GluTrpArgAspProGlyMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln 310  
Qy 778 GGGTTCAAGAACTCCCGACCATCTTTGACGAGCCCTCACAGACCTGGCCCACTTC 837  
Db GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaGlyPhe 330  
Qy 838 AGGATCAACAACCTCAGTGGACCTCTCCTCAGTACGTGGATGACCTGCTTCTCGCGGA 897  
Db ArgIleGlnHisProAspLeuLeuLeuGlnTyrValAspAspLeuLeuLeuAlaAla 350  
Qy 898 GCCACCAACAGGACTCTTAGAAGGCACGAAGCACTACTGCTGGAAATTTGCTGACCTA 957  
Db ThrSerGluLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAspLeu 370  
Qy 958 GGCTACAGAGCTCTGCTAAGAGGCCCGACATTTGACGAGAGAGAGATACTCTGGGG 1017  
Db GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly 390  
Qy 1018 TACAGTTTGGGGACGGGACGATGGCTGACGAGGACGAGCAAGAAACAACTGTAGTCCAG 1077  
Db TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGluThrValMetGly 410  
Qy 1078 ATACCGGCCCCAACCAACAGCCAAACAAATGAGAGAGATTTTGGGACAGCTGGATTTCG 1137  
Db GlnProIleProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys 430  
Qy 1138 AGACTGTGGATCCCGGGTTTGGACCTTAGCAGCCCACTCTACCCGCTAACCAAGAA 1197  
Db ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr 450  
Qy 1198 AAAGGGAAATTTCTCTGGGCTCTCTGAGCACACAGAGCATTTGATGTATCAAAAGGCC 1257  
Db GlyThrLeuPheAsnTrpGlyProAspGlnLysAlaTyrGlnGluIleLysGlnAla 470  
Qy 1258 CTGCTGACGCGACCTGCTCTGGCCCTCCCTGACCTAAACCTTTTACCCCTTTATGTG 1317  
Db LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluPheVal 490  
Qy 1318 GATGAGGTAAGGAGTAGCCCGGGAGTTTAAACCAACCCCTAGGACCACTGGAGAGA 1377  
Db AspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg 510  
Qy 1378 CCTCTGCTACCTGTCAAGAAGCTCGATCCTCTAGCAGTGGTTGGCCCATATGCTG 1437  
Db SerValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTyrProCysLeu 530  
Qy 1438 AAGGCTATCGAGCTGTGGGCCATATCTGTCAGGACCGCTGACAAATTCAGCTTTGGACAG 1497  
Db ArgMetValAlaAlaIleAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGln 550





Db 551 ProLeuValIleLeuAlaProHisAlaValGluAlaLeuValLysGlnProProAspArg 570  
Qy 1558 TGGATGACCAACGCCCGCATGACCCACTATCAAGCCTGCTTC- --ACAGAGAGGGTC 1614  
Db 571 TrpLeuSerAsnAlaAargMetThrHisTyrglnAlaMetLeuLeuAspThrAspArgVal 590  
Qy 1615 ACGTCGCTCCACAGCGCTCTCAACCTGCCACTCTTCCTGCTGAAGAGACTGTAGAA 1674  
Db 591 GlnPheGlyProValValAlaLeuAsnProAlaThrLeuLeuPro- --LeuProGluGlu 609  
Qy 1675 CCAGTGACTCATGATTGGCCATCACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTT 1734  
Db 610 GlyAlaProHisAspCysLeuGluLeuAlaGluThrHisGlyThrArgProAspLeu 629  
Qy 1735 ACAGACATACCGCTGACTGGAGAAAGTGTAACTGGTTCACTGACGGAGAGCAGCTATGTG 1794  
Db 630 ThrAspGlnProIleProAspAlaAspHisThrTrpTyrrThrAspGlySerSerPheLeu 649  
Qy 1795 GTGGNAGTAAAGAGATGGCTGGGGCGCGGTGGTGGACGGGACCGCACCATCTGGGCC 1854  
Db 650 GlnGluGlyGlnArgLysAlaGlyAlaValThrThrGluThrGluValIleTrpAla 669  
Qy 1855 AGCAGCCTGCCGGAAGGAATTCAGCACAAAGGCTGAGTCATGCGCCCTCACCAAGCT 1914  
Db 670 ArgAlaLeuProAlaGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAla 689  
Qy 1915 TTGGCGCTGGCGGAAGGAATTCATAACAACTTTATACGACAGCAGGTATGCTTTGCG 1974  
Db 690 LeuLysMetAlaGluGlyLysArgLeuAsnValTyrrThrAspSerGlnTyrrAlaPheAla 709  
Qy 1975 ACTGCACATGATGGGGCCCATCTATAACAAAGGGGTGCTTACTCAGCAGGGAGG 2034  
Db 710 ThrAlaHisIleHisGlyGluIleTyrrArgArgGlyLeuLeuThrSerGluGlyArg 729  
Qy 2035 GAATAAAGAACAAAGGAATTCCTAAGCCTATTAGACCGGTACATTACCAAAAGG 2094  
Db 730 GluIleLysAsnLysSerGluIleLeuAlaLeuLysAlaLeuPheLeuProLysArg 749  
Qy 2095 CTAGCTATTATACACTGCTCGGACATCAGAAAGCTCAATCATCTCATATCCAGAGGAAC 2154  
Db 750 LeuSerIleIleHisCysLeuGlyHisGlnLysGlyAspGlyAlaGluAlaAargGlyAsn 769  
Qy 2155 CAGATGCTACCGGGTTCACACGAGCAGCC-----CAGGGTGTAACTT 2202  
Db 770 ArgLeuAlaAspGlnAlaAlaAargGluAlaIleLysThrProProAspThrSerThr 789  
Qy 2203 CTGCTATAATAGAAATGCCAAAGCCCGAGACCCAGACAGTACACCTCAGAGAC 2262  
Db 790 LeuLeuIleGluAspSerThrProTyrrThrProAlaTyrrPheHisTyrrThrGluThrAsp 809  
Qy 2263 TGGCAAGAGATAAAAAAGATAGACCACTTCTCTGAGACTCCGGAGGGGACCTGCTATACC 2322  
Db 810 LeuLysLysLeuArgGluLeuGlyAlaThrTyrrAsnGlnSerLysGly---TyrrTrpVal 828  
Qy 2323 TCAGATGGGAAGGAATCTCTGCCCCCAAGAGGGTTAGAAATATGTCACACAGATACAT 2382  
Db 829 PheGlnGlyLysProValMetProAspGlnPheValPheGluLeuLeuAspSerLeuHis 848  
Qy 2383 CGTCTAACCCCTAGAACTAAACACCTCGACGAGTGTGTC-----AGAACATCCCT 2436  
Db 849 ArgLeuThrHisLeuGlyTyrrGlnLysMetLysAlaLeuLeuAspArgGlyGluSerPro 868  
Qy 2437 TATCATGTTCTGAGGCTACCCAGGAGTGGCTGACTCGGTGCTCAAAACATTGTGCTCGC 2496  
Db 869 TyrrTyrrMetLeuAsnArgAspLysThrLeuGlnTyrrValAlaAspSerCysThrValCys 888  
Qy 2497 CAGCTGTTTAACTGCTTAATCTTCCAGAAATGCTTCCAGGGGAAGAGACTTAAGGGGAAGCCAC 2556  
Db 889 AlaGlnValAsnAlaSerLysAlaIleGlyAlaGlyValArgValArgGlyHisArg 908  
Qy 2557 CAGGGCGCTCCTGGGAGTGGACTTCACGTAGGTAAAGCGCGCTAAATACGGAAACAA 2616  
Db 909 ProGlySerHisTrpGluIleAspPheThrGluValLysProGlyLeuTyrrTyrrLys 928

Qy 2617 TACCTATTGGTTTGTAGACACCTTTTACGATGGGTAGAGGCTTATCTCTACTAAGAAA 2676  
Db 929 TyrLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaPheProThrLysArg 948  
Qy 2677 GAGACTTCAACCGTGGTGGCTAAATAATACTGGAAGAAATTTTCCAAAGATTTTGAATA 2736  
Db 949 GluThrAlaArgValSerLysLysLeuLeuGluIlePheProArgPheGlyMet 968  
Qy 2737 CCTAAGGTAAATAGGTCCAGCAATGGTCCAGCTTTTGTGTCAGGTAAAGTCAGGACTG 2796  
Db 969 ProGlnValLeuGlySerAspAsnGlyProAlaPheThrSerGlnValSerGlnSerVal 988  
Qy 2797 GCCAAGATATTGGGATTTGATTGGAACTGCAATTGTGCATACACAGCCCAAGCTCAGGA 2856  
Db 989 AlaAspLeuLeuGlyIleAspTrpLysLeuHisCysAlaTyrrArgProGlnSerSerGly 1008  
Qy 2857 CAGTAGAGAGGATGAATAGAACCATTAAGAGACCTTACTAAATGAGC- --GCGGAG 2913  
Db 1009 GlnValGluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuAlaAla 1028  
Qy 2914 ACTGGCCTTAATGATTGATAGTCTCTGCTGCCCTTGTGCTTTTGGGTTAGGAACACC 2973  
Db 1029 ProGlyThrArgAspTrpValLeuLeuLeuProLeuAlaLeuTyrrArgAlaAargAsnThr 1048  
Qy 2974 CCTGGACAGTTTGGGCTGACCCCTATGAATTACTCTACGGGGGACCCGCCCACTTGTA 3033  
Db 1049 ProGlyProHisGlyLeuThrProTyrrGluIleLeuTyrrArgAlaProProLeuVal 1068  
Qy 3034 GAAATTCTCTGTACATAGTGTGCTGCTGCTTTTCCAGCGCTTTGTTCTCTAGGCTC 3093  
Db 1069 AsnPheHisAspProAspMetSerGluLeuThrAsnSerProSerLeuGlnAlaHisLeu 1088  
Qy 3094 AAGCACTTCAGTGGGTGACACACAGCGCTGAGGCAACTCCGGGAGGCTTACTCAGGA 3153  
Db 1089 GlnAlaLeuGlnThrValGlnArgGluIleTrpAsnProLeuAlaGluAlaTyrrArgAsp 1108  
Qy 3154 GAGAGGACATTCGAG- --ATCCACACATCGTTTCCAGTGGGAGATTTCAGTCTACGTAGA 3210  
Db 1109 GlnLeuAspGlnProValIleProHisPheArgIleGlyAspSerValTrpValArg 1128  
Qy 3211 CGCACCGTCGACGAAACCTCGAGACTCGGTGGAAGGCGCTTATCTCGTACTTTTGACC 3270  
Db 1129 ArgHisGlnThrLysAsnLeuGluProArgTrpLysGlyProTyrrThrValLeuLeuThr 1148  
Qy 3271 ACACCAACGCTGTGAAGTCGAAGNATCTCCACCTGGATCCATGCATCCACGTTAAA 3330  
Db 1149 ThrProThrAlaLeuLysValAspGlyIleSerAlaTrpIleHisAlaHisValLys 1168  
Qy 3331 CCGGCG- ----CCACCTCCCGATTCCGGGTGGAAGCCGAAAGACTGAAAAATCCCTT 3384  
Db 1169 AlaAlaThrThrProProLysProSerTrpArgValGlnArgSerGlnAsnProLeu 1188  
Qy 3385 AAGCTTTCGCTCCATCGCTGGTTCTCCT 3411  
Db 1189 LysIleArgLeuThrArgGlyAlaPro 1197  
RESULT 8  
ADH76470  
ID ADH76470 standard; protein; 1199 AA.  
XX  
AC ADH76470;  
XX  
DT 15-APR-2004 (first entry)  
XX Murine leukemia virus pol gene derived protein.  
DE chimeric plasmid; replicative retroviral genome; gag; pol;  
XX murine leukemia virus; MLV; env; gibbon ape leukemia virus; GalV; virion;  
KW MLV-GalV-type; gene therapy; pAM.  
XX Murine leukemia virus.  
OS  
XX

PN FR2832424-A1.  
 XX 23-MAY-2003.  
 XX 20-NOV-2001; 2001FR-00014976.  
 XX 20-NOV-2001; 2001FR-00014976.  
 PR (GENE-) GENETHON III.  
 XX Audit M, Cosset FL;  
 XX WPI; 2003-471779/45.  
 XX Chimeric plasmid containing replicative retroviral genome, useful for  
 PT making positive control virions in testing for replication-competent  
 PT retrovirus.  
 XX Claim 4; SEQ ID NO 2; 70pp; French.  
 XX The invention relates to a novel chimeric plasmid comprising a  
 CC replicative retroviral genome. The replicative retroviral genome  
 CC comprises: the gag and pol sequences from a murine leukemia virus (MLV);  
 CC and a chimeric env sequence comprising regions corresponding to parts of  
 CC the envelope derived from: an MLV genome; and a gibbon ape leukemia virus  
 CC (GALV). Virions produced by expressing the viral genome of the chimeric  
 CC plasmid are useful as positive controls in a test for detection of  
 CC replication-competent retroviruses in preparations of MLV-GALV-type  
 CC retroviral vectors. For example, to ensure that the MLV-GALV-type  
 CC retroviral vectors, intended for gene therapy, have no capacity for  
 CC replication. This sequence represents the protein derived from an MLV pol  
 CC gene from a pAM plasmid of the invention.  
 XX  
 SQ Sequence 1199 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 1199  
 Score: 2808.50 Matches: 722  
 Percent Similarity: 76.2% Conservative: 160  
 Best Local Similarity: 62.3% Mismatches: 245  
 Query Match: 61.6% Indels: 31  
 DB: 7 Gaps: 11

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x ADH76470 (1-1199)

QY 4 GGTGCCACAGGCAACACAGATATCCATGGACTACCCGAGAACAGTTGACTTGGGATG 63  
 Db 51 GlyAlaThrGlyGlyLysArgTyrArgTrpThrThrAspArgLysValHisLeuAlaThr 70  
 QY 64 GGACGGGTAAACCACTCGTCTTCTGTCATACCTGAGTCCCGACACCCCTCTTAGGTAGA 123  
 Db 71 GlyLysValThrHisSerPheLeuHisValProAspCysProTyrProLeuLeuGlyArg 90  
 QY 124 GACTTATTGACCAAGATGGGAGCAACAAATTTCTTTTGA---CAAGGGAAACAGAAAGT 180  
 Db 91 AspLeuLeuThrLysLeuLysAlaGlnIleHisPheGluGlySerGlyAlaGlnValMet 110  
 QY 181 TCTGCAATAACAAACCTATCACTCTGTGTGACCCCTCCCAATTAGATGACGAATATCGACTA 240  
 Db 111 GlyProMetGlyGlnProLeuGlnValLeuThrLeuAenIleGluAspGluHisArgLeu 130  
 QY 241 TACTCTCCCTAGTAGTAAGCCTGATCAAAATATA---CAATCTGTTGGACAGTTTCCC 297  
 Db 131 HisGluThrSerLysGluProAspValSerLeuGlySerThrTrpLeuSerAspPhePro 150  
 QY 298 CAAGCCTGGGCAAAACCGAGGATGGTTGGCAAAAGCAAGTTCCCCCAACAGTTATT 357  
 Db 151 GlnAlaTrpAlaGluThrGlyGlyMetGlyLeuAlaValArgGlnAlaProLeuIle 170  
 QY 358 CAACATGAGGCGCAGTCCACACCATGTGTCAGTCAGTACAGTACCCCTTGAGTAAAGAGCT 417  
 Db 171 ProLeuLysAlaThrSerThrProValSerIleLysGlnTyrProMetSerGlnGluAla 190

QY 418 CAAGAAGGAATTCGGCCGCATGTCCAAGATTAATCCAACAGGGCATCTAGTTCCTGTC 477  
 Db 191 ArgLeuGlyIleLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys 210  
 QY 478 CAATCTCCCTGGAATACCTCCCTGCTACCGGTTAGAAAGCCTGGAGCTAATAGTATCGA 537  
 Db 211 GlnSerProTrpAenThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg 230  
 QY 538 CCAGTACAGGACTTGAGAGAGGTCAATAAAGGGTGCAGGATATACACCCACAGTCCCG 597  
 Db 231 ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro 250  
 QY 598 AACCTTTATAACCTCTTGTGTCTCTCCACCCCAACCGAGCTCGTATACAGTATTGGAC 657  
 Db 251 AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp 270  
 QY 658 TTAAGGATGCTTCTTCTGCTGAGATTACACCCCACTAGCCCAACCACTTTTGGCTTC 717  
 Db 271 LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPhe 290  
 QY 718 GAATGGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGAGCCCGACTGCCCAA 777  
 Db 291 GluTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln 310  
 QY 778 GGGTTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACTCGCCCAACTTC 837  
 Db 311 GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe 330  
 QY 838 AGGATCCAACACCTCAGGTGACCTCTCCTCAGTACGTGGATGACCTGCTTCTGGCGGA 897  
 Db 331 ArgIleGlnHisProAspLeuIleLeuLeuGlnTyrValAspAspLeuLeuLeuAlaAla 350  
 QY 898 GCCACCAACAGGACTCTTAGAAGCGACGAGCACTACTGCTGGAATTCCTGACCTA 957  
 Db 351 ThrSerGluLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAsnLeu 370  
 QY 958 GGCTACAGAGCTCTGTAGAAGCGCCAGATTTCAGAGAGAGAGTAGTAACATCTTGGG 1017  
 Db 371 GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly 390  
 QY 1018 TACAGTTTGGCGGACGGCGAGCGATGCTGACGAGGACGCGAAGAAACCTGTAGTCAG 1077  
 Db 391 TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGluThrValMetGly 410  
 QY 1078 ATACGGCCCCCAACCAACAGCCAAACAAATGAGAGAGTTTTCGGGACAGCTGGATTTC 1137  
 Db 411 GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys 430  
 QY 1138 AGACTGTGGATCCCGGGTTTCGACCTTAGCAGCCCCACCTCTACCCGCTAACCAAGAA 1197  
 Db 431 ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr 450  
 QY 1198 AAAGGGGAATTCCTCGGCTCTGAGCACACGAAAGCAATTTGATGCTATCAAAAGGCC 1257  
 Db 451 GlyThrLeuPheAsnTrpGlyProAspGlnLysAlaTyrGlnGluIleLysGlnAla 470  
 QY 1258 CTGCTGACGCGACCTGCTCTGCGCCCTCCCTGACGTAACCTTAACCTTTACCTTATGTC 1317  
 Db 471 LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal 490  
 QY 1318 GATGAGCTAAGGAGTAGCCCGGGAGTTTAAACCCCAACCCCTAGGACCATGAGAGAAGA 1377  
 Db 491 AspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg 510  
 QY 1378 CCTGTGCTCTACCTGTCAAAGAAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCTGT 1437  
 Db 511 ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTrpProProCysLeu 530  
 QY 1438 AAGGCTATCGAGCTGGGCCATACTGTCAGGACGCTGACAAATTCACCTTTGGGACAG 1497  
 Db 531 ArgMetValAlaAlaIleAlaValLeuThrLysAspAlaGlyLysLeuThrMetGln 550  
 QY 1498 AATATACTGTAATAGCCCCCCCCCATGTCATTTGGAGAACATCTGTTGGCAGCCCCCAGACCGA 1557



PN	W0200042199-A1.	QY	358	CAACTGAAGCCAGTCCACACAGTGTCTCAGTCAGACAGTACCCCTTGTAGTAAGAAGCT	417
XX					
PD	20-JUL-2000.	Db	171	ProLeuLysAlaThrSerThrProValSerIleLysGlnTyrProMetSerGlnGluAla	190
XX					
XX	14-JAN-2000; 2000WO-US000896.	QY	418	CAAGAAGGAATTCGGCCGATGTCCTCAAGATTAATCAACAGGGCATCTAGTTCCTGTC	477
XX					
XX	15-JAN-1999; 99US-0116099P.	Db	191	ArgLeuGlyIleLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys	210
XX					
PA	(MOLE-) MOLECULAR BIOLOGY RESOURCES.	QY	478	CAATCTCCCTGGATATCTCCCTGCTACCGTTAGAAAGCTGGGACTATGACTATCGA	537
XX					
XX	Swaminathan N;	Db	211	GlnSerProTrpAsnThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg	230
XX					
DR	WPI; 2000-482830/42.	QY	538	CCAGTACAGGACTTGAGAGAGGTCAATAAAGCGGTGCAGGATATACACCAACAGTCCCG	597
XX					
XX	Novel genes encoding reverse transcriptase polypeptides modified by	Db	231	ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro	250
PT	altering or adding the integrase domains by truncation internally and/or				
PT	at the C-termini, useful in cDNA synthesis and amplification procedures.	QY	598	AACCTTTAATACCTCTTGTGTCTCTCCACCCCAACGGAGCTGGTATACAGTATTGGAC	657
XX					
XX	Claim 1; Page 154-159; 189pp; English.	Db	251	AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp	270
XX					
CC	This invention relates to isolated polynucleotide sequences encoding a	QY	658	TTAAAGGATCCCTTCTCTGCTGAGATTACCCCACTAGCCCAACCACTTTTTCCTTC	717
CC	polypeptide with RNA dependent DNA polymerase (reverse transcriptase RT)				
CC	activity. RTs are found in a variety of retroviruses and their defining	Db	271	LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPhe	290
CC	activity is the ability to synthesise a cDNA strand using an RNA				
CC	template. The invention includes sequences AA87808-A87822 and AA87840-	QY	718	GAATGGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTCCCAA	777
CC	A87842 which represent nucleotide sequences encoding RT polypeptides.				
CC	Included in the invention are reverse transcriptase protein sequences	Db	291	GluTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln	310
CC	AA812989-B12995. Sequences AA87829-A87839 and AA87843-A87897 represent	QY	778	GGGTTCAGAACTCCCGACCATCTTTGACGAAGCCCTCACAGAGACCTGGCCCACTTC	837
CC	oligonucleotides used in the identification and synthesis of the RT				
CC	nucleotide sequences of the invention. Many of the RT nucleotide	Db	311	GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe	330
CC	sequences encode modified RT proteins, which exhibit improved stability				
CC	and/or improved solubility, relative to naturally occurring reverse	QY	838	AGATTCACACCTCCTAGGTGACCTCTCCTCAGTACCTGATGACCTCTCTCGCGGA	897
CC	transcriptases. The modified RT nucleotide and protein sequence are used				
CC	in improved polymerase chain reaction (PCR) methods. The nucleotide	Db	331	ArgIleGlnHisProAspLeuLeuLeuGlnTyrValAspAspLeuLeuLeuAlaAla	350
CC	sequences can be used in sequencing methods. The present sequence				
CC	represents a reverse transcriptase protein of the invention	QY	898	GCACCAACACAGGACTCTTAGAAGCACCAAGCACTACTCTGCTGGAATTTCTGTGACCTA	957
XX					
SQ	Sequence 1199 AA;	Db	351	ThrSerGluLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAsnLeu	370
	Alignment Scores:	QY	958	GGCTACAGAGCTCTGCTAAGAAGCCAGATTTCGAGGAGAGAGAGTAACATCTGGGG	1017
	Pred. No. : 0 Length: 1199				
	Score: 719 Matches: 719	Db	371	GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly	390
	Percent Similarity: 76.3% Conservative: 165				
	Best Local Similarity: 62.0% Mismatches: 242	QY	1018	TACAGTTTCGGGACGGCGACCGATGCTGACGGGACGACGAGAAACCTGAGTCCAG	1077
	Query Match: 61.4% Indels: 33				
	DB: 3 Gaps: 12	Db	391	TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGluThrValMetGly	410
	US-10-723-552-3_COPY_2307_5741 (1-3435) x AAB12994 (1-1199)	QY	1078	ATACCGGCCCCAACACACAGCCAAATGAGAGAGTTTTTTGGGACAGCTGGATTTTGC	1137
		Db	411	GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys	430
		QY	1138	AGACTGTGATTCGCGGGTTTGCACCTTACGACGCCCTCTCTACCCGCTAACCAAGAA	1197
		Db	431	ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr	450
		QY	1198	AAAGGGGAATTCCTCTGGCTCCTGACACACAGAGGCAATTTGATGCTATCAAAAGGCC	1257
		Db	451	GlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTyrGlnGluIleLysGlnAla	470
		QY	1258	CTGCTGACGCGACCTGCTCTGGCCCTCCTGACCTAACTAAACCTTTTACCTTTATGTG	1317
		Db	471	LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal	490
		QY	1318	GATGAGCGTAAGGAGTAGCCCGGGGATTTTAAACCAAAACCTTAGGACCATGAGAGA	1377
		Db	491	AspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg	510
		QY	1378	CCTGCTCCCTACCTGTCAGGAAGCTCGCTCCTAGCCAGTGTGTGGCCCATATGCGCTG	1437
		Db	511	ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTyrProProCysLeu	530
		QY	1438	AAGGCTATCGCAGCTGTGGCCATCTGCTCAAGGACGCTGCAAAATTGACTTTGGGACAG	1497



```

Db 531 ArgMetValAlaAlaAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGln 550
Qy 1498 AATATAACATGAATAGCCCCCATCGATTTGGAGAACATCGTTTGGCAGCCCCCAGACCGA 1557
Db 551 ProLeuValIleLeuAlaProHisAlaValGluAlaLeuValLysGlnProAspArg 570
Qy 1558 TGGATGACCAACGCCCGCATGACCCACTATCAAGCCCTGCTCTC---ACAGAGAGGGTC 1614
Db 571 TrpLeuSerAsnAlaArgMetThrHisTyrGlnAlaLeuLeuLeuLeuAspThrAspArgVal 590
Qy 1615 ACGTTGCTCCACAGCGCTCTCAACCTCGCACTCTTCCTGCTGGAAGAGACTGATGAA 1674
Db 591 GlnPheGlyProValValAlaLeuAsnProAlaThrLeuLeuPro---LeuProGluGlu 609
Qy 1675 CCAGTGACTCATGATTGGCCATCACTATTGATTGAGGAGACTGGGCTCGCAAGGACCTT 1734
Db 610 GlyLeuGlnHisAsnCysLeuAspIleLeuAlaGluAlaHisGlyThrArgProAspLeu 629
Qy 1735 ACAGACATACCGCTGACTGGAGAAAGTGTAACTGGTTTCACTGACGGAGCAGCTATGTG 1794
Db 630 ThrAspGlnProLeuProAspAlaAspHisThrTrpTyrThrAspGlySerSerLeuLeu 649
Qy 1795 GTGNAAGGTAGAGATCGCTGGCGCGCGTGTGGACGGGACCGCACCATCTGGGCC 1854
Db 650 GlnGluGlyGlnArgLysAlaGlyAlaValThrThrGluThrGluValIleTrpAla 669
Qy 1855 AGCAGCTGCGCGAAGGAACCTTCAGCACAAAGCTGAGTCATGCGCCCTCACCAAGCT 1914
Db 670 LysAlaLeuProAlaGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAla 689
Qy 1915 TTGGCGCTGGCGAAGGAATTCATAACATTTATACGACAGCAGCTATGCTTTGCG 1974
Db 690 LeuLysMetAlaGluGlyLysLysLeuAsnValTyrThrAspSerArgTyrAlaPheAla 709
Qy 1975 ACTGCACAGTACATGGGGCCATCTATAACAAGGGGGTGTCTTACTTCAGCAGGGAGG 2034
Db 710 ThrAlaHisIleHisGlyGlyGlyLeuIleTyrArgArgGlyLeuLeuThrSerGluGlyLys 729
Qy 2035 GAAATAAGAACAAAGAGGAATTTCTAAGCCTATTAGAACCGGTACATTTTACCAAAAGG 2094
Db 730 GluIleLysAsnLysAspGluIleLeuAlaLeuLysAlaLeuPheLeuProLysArg 749
Qy 2095 CTAGCTATTATACATGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAC 2154
Db 750 LeuSerIleIleHisCysProGlyHisGlnLysGlyHisSerAlaGluAlaArgGlyAsn 769
Qy 2155 CAGATGCTGACCGGGTTCACCAAGCAGCGCCAGGGTGTAACTTCTGCTCTATATA 2214
Db 770 ArgMetAlaAspGlnAlaAlaArgLysAlaAla-----IleThr 782
Qy 2215 GAAATGCCCAAGCC-----CCAGAACCCAGA 2241
Db 783 GluThrProAspThrSerThrLeuLeuIleGluAsnSerSerProTyrThrSerGluHis 802
Qy 2242 CGACAGTACACCTCAGAGACTGGCAAGAGATAAAGAGATAGACCAG---TTCCTGAG 2298
Db 803 PheHisTyrThrValThrAspIleLysAspLeuThrLysLeuGlyAlaIleTyrAspLys 822
Qy 2299 ACTCCGAGAGGACCTGCTATACCTCAGATGGGAAGGAATCTTCGCCCCCAAGAGGG 2358
Db 823 ThrLysLysTyrTrpValTyr-----GlnGlyLysProValMetProAspGlnPheThr 840
Qy 2359 TTAGAATATGCCAACAGATACATCGTCTAACCCCTAGCACTAGAACTAAACCTGCAGCAG 2418
Db 841 PheGluLeuLeuAspPhePheLeuHisGlnLeuThrHisLeuSerPheSerLysMetLysAla 860
Qy 2419 TTGCTCAGAAACA-----TCCCTTATCATGTTCTCTGAGGCTTACCAGGAGTGGCTGACTCG 2472
Db 861 LeuLeuGluArgSerHisSerProTyrTyrMetLeuAsnArgAspArgThrLeuLysAsn 880
Qy 2473 GTGCTCAACATTTGTGTGGCCCTGCCAGCTGTTAATCTTAATCTTCCAGATGCCCTCCA 2532

```

```

Db 881 IleThrGluThrCysLysAlaCysAlaGlnValAsnAlaSerLysSerAlaValLysGln 900
Qy 2533 GGGAGAGACTAAGCGGGAAGCCACCCAGCGCTCACTGGAGTGGACTTCACTCAGAGTA 2592
Db 901 GlyThrArgValArgGlyHisArgProGlyThrHisTrpGluIleAspPheThrGluIle 920
Qy 2593 AAGCCGGCTAAATACGGAACAAATACCTATTGGTTTTTTGTAGACACCTTTTCAGATGG 2652
Db 921 LysProGlyLeuTyrGlyTyrLysTyrLeuLeuValPheIleAspThrPheSerGlyTrp 940
Qy 2653 GTAGAGCTTATCTTACTTAAGAAAGAGACTTCAACCGTGGTGGTGGTAAAAAATACTGAA 2712
Db 941 IleGluAlaPheProThrLysLysGluThrAlaLysValValThrLysLysLeuLeuGlu 960
Qy 2713 GAAATTTTCCAGATTGGAAATACCTAAGGTAAATAGGTTCAGACAATAGTGTCCACTTTT 2772
Db 961 GluIlePheProArgPheGlyMetProGlnValLeuGlyThrAspAsnGlyProAlaPhe 980
Qy 2773 GTTCCCGAGGTAACTCAGGGACTCGCCAAGATATTGGGGATTGATTGGAAACTGCAATTGT 2832
Db 981 ValSerLysValSerGlnThrValAlaAspLeuLeuGlyIleAspTrpLysLeuHisCys 1000
Qy 2833 GCATACAGACCCCAAGCTCAGACAGTAGAGAGATGAATAGAACCATTAAGAGACC 2892
Db 1001 AlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThr 1020
Qy 2893 CTTACTAAATTGACCGCGAGACTGGCGTTAATGATTGGATAGCTCTCCCTCCCTTTGTG 2952
Db 1021 LeuThrLysLeuThrLeuAlaThrGlySerArgAspTrpValLeuLeuLeuProLeuAla 1040
Qy 2953 CTTTATAGGTTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCTATGAATTAATCTTAC 3012
Db 1041 LeuTyrArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGluIleTyr 1060
Qy 3013 GGGGGACCCCGCCCATTTGGTAGAAATTTGCTTCTGTACATAGTGTGCTGCTGCTTTCC 3072
Db 1061 GlyAlaProProProLeuValAsnPheProAspProAspMetThrArgValThrAsnSer 1080
Qy 3073 CAGCCTTTGTTCTCTAGGCTCAAGGCACCTTGAGTGGGTGAGACAACGAGCGTGGAGGCA 3132
Db 1081 ProSerLeuGlnAlaHisLeuGlnAlaLeuTyrLeuValGlnHisGluValTrpArgPro 1100
Qy 3133 CTCGGGAGCGCTACTCAGGAGGAGGAGACTTGCAG---ATCCCACTATCGTTTCCAGTG 3189
Db 1101 LeuAlaAlaIleTyrGlnGluGlnLeuAspArgProValValProHisProTyrArgVal 1120
Qy 3190 GGAGATTCACTCTAGTTAGACCGCCCGTGGAGAAACCTCGAGACTCGGTGGAAGGGC 3249
Db 1121 GlyAspThrValTrpValArgHisGlnThrLysAsnLeuGluProArgTrpLysGly 1140
Qy 3250 CCTTATCTCTCTTTTTCACCAACCAACCGCTGTGAAGTCCGAAGGAATCTCCACCTGG 3309
Db 1141 ProTyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleAlaAlaTrp 1160
Qy 3310 ATCCATGCACTCCACGTTAAACCGCGCCACCT-----CCCGATTGG-----GGG 3354
Db 1161 IleHisAlaAlaHisValLysAlaAlaAspProGlyGlyGlyProSerSerArgLeuThr 1180
Qy 3355 TGGAAAGCCGAAAGACTGAAATTCCTTTAAGCTTCGCTCCATCCGCTGGTTCT 3411
Db 1181 TrpArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGluAlaPro 1199

```

## RESULT 10

AAB10044

ID AAB10044 standard; protein; 1737 AA.

XX

AC AAB10044;

XX

DT 02-NOV-2000 (first entry)

XX

DE MMLV gag-pol protein.

XX

KW Glycoprotein; gag gene; pol gene; GP-1; GP-2; anti-HIV; cytosstatic;

KW gene therapy; treatment; infectious disease; HIV; AIDS; neoplasm;  
KW Carcinoma; melanoma.  
XX  
XX Moloney murine leukemia virus.  
XX EP1006196-A2.  
XX  
XX  
XX PD 07-JUN-2000.  
XX  
XX PF 25-NOV-1999; 99EP-00250415.  
XX PR 26-NOV-1998; 98DE-01056463.  
XX (PETT-) PETTE INST HEINRICH.  
XX  
XX PI Von Laer MD;  
XX  
XX DR WPI; 2000-378268/33.  
XX  
XX PT New retroviral packing cell useful as pharmaceutical carrier in gene  
XX therapy for treatment of HIV and neoplasms, comprises retroviral genes  
XX and glycoproteins.  
XX  
XX PS Disclosure; Page 32-36; 69pp; German.  
XX  
XX CC This invention describes a novel retroviral packing cell (I), comprising  
XX the retroviral genes gag, pol and glycoproteins GP-1 and GP-2 of the LCMV  
XX coding gene gp, or a part of these. The products of the invention have  
XX anti-HIV and cytostatic activity and can be used for gene therapy. (I) is  
XX useful for in vitro infection of cells, especially hematopoietic stem  
XX cells, for expression of transgenes in cells and as a pharmaceutical  
XX carrier for gene therapy. (I) is therefore useful in the treatment of  
XX infectious diseases (HIV or AIDS), neoplasms (carcinomas), melanomas and  
XX other diseases. This sequence represents the Moloney murine leukemia  
XX virus (MMuLV) gag-pol protein which is described in the method of the  
XX invention  
XX  
XX SQ Sequence 1737 AA;  
  
Alignment Scores:  
Pred. No.: 0 Length: 1737  
Score: 3797.50 Matches: 719  
Percent Similarity: 76.3% Conservative: 165  
Best Local Similarity: 62.0% Mismatches: 242  
Query Match: 61.4% Indels: 33  
DB: 3 Gaps: 12  
  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x AAB10044 (1-1737)  
  
QY 4 GGTGCCACGGGCAACACAGTATCCATGGACTACCCGAGCAACAGTTGACTTGGAGTG 63  
Db GlyAlaThrGlyGlyLysArgTyrArgTrpThrAspArgLysValHisLeuAlaThr 608  
  
QY 64 GGACGGGTAAACCACTCTGCTGTCATCTAGTGCACCGAGCCACCTCTTAGGTAGA 123  
Db GlyLysValThrHisSerPheLeuHisValProAspCysProTyrProleuLeuGlyArg 628  
  
QY 124 GACTTATTGACCAAGATGGGACGACAAATTTCTTTTGA---CAAGGAAACCAAGAGTG 180  
Db AspLeuLeuThrLysLeuLysAlaGlnHisPheGluGlySerGlyAlaGlnValMet 648  
  
QY 181 TCTGCAATAACAAACCTATCACTGTGTGACCTCCATAGATGAGCAATATCGACTA 240  
Db GlyProMetGlyGlnProLeuGlnValLeuThrLeuAsnIleGluAspGluHisArgLeu 668  
  
QY 241 TACTCTCCCTAGTAAAGCCTGATCAAAATATA---CAATTCTGGTTGGAAACAGTTTCCC 297  
Db HisGluThrSerLysGluProAspValSerLeuGlySerThrTrpLeuSerAspPhePro 688  
  
QY 298 CAAGCTTCGGGAGAAACCGCAGGATGGTTGGCAAGCAAGTTCCCCCAAGATTATT 357  
Db GlnAlaTrpAlaGluThrGlyGlyMetGlyLeuAlaValArgGlnAlaProleuIle 708

QY 358 CAACTGAAGGCCAGTGCACACACAGTGTCACTCAGACAGTACCCTTGAGTAAAGAAGCT 417  
Db ProLeuLysAlaThrSerThrProValSerIleLysGlnTyrProMetSerGlnGluAla 728  
  
QY 418 CAAGAAGGAATTCGGCCGCATGTCACAAAGATTAATCCAACACAGGCGCATCTAGTTCTCTGTC 477  
Db ArgLeuGlyIleLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys 748  
  
QY 478 CAATCTCCCTGGAATACTCCCTGCTACCGGTTAGAAAGCCTGGAGCTAATGACTATCGA 537  
Db GlnSerProTrpAsnThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg 768  
  
QY 538 CCAGTACAGGACTTGAGAGAGTCAATAAAGCGGTGCAGGATATACACCCACAGTCCCG 597  
Db ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro 788  
  
QY 598 AACCCCTTATACCTCTTTGTGTCTCTCCACCCCAACCGAGCTGGTATACAGTATTGGAC 657  
Db AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp 808  
  
QY 658 TTAAGGATGCTCTTCTGCTGAGATTACACCCCTAGCCACCACTTTTTCCTTC 717  
Db LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPhe 828  
  
QY 718 GAATGGAGAGATCAGGTACGGGAAGAACCGGCGAGCTCACCTGACCCGACTGCCCAA 777  
Db GluTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln 848  
  
QY 778 GGGTTCGAAGAACTCCCGACCATCTTTGACGAAGCCCTTACACAGAGACCTGGCCAACTTC 837  
Db GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe 868  
  
QY 838 AGGATCCAAACCTCCTAGGTGACCTCTCTCAGTACGTGGATGACCTGCTTCTGGCGGA 897  
Db ArgIleGlnHisProAspLeuIleLeuLeuGlnTyrValAspAspLeuLeuAlaAla 888  
  
QY 898 GCCACCAACAGGACTGCTTAGAAGGACGACGAGCTACTGCTGGAATTCCTGACCTA 957  
Db ThrSerGluLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAsnLeu 908  
  
QY 958 GGCTACAGAGCTCTGCTAAGAAGCCGAGATTTGAGGAGAGAGGTAACTACTTGGGG 1017  
Db GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly 928  
  
QY 1018 TACAGTTTGGGAGCGGACGAGTGTGCTGACGAGGACCGAAGAAACTGCTAGTCCAG 1077  
Db TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGluThrValMetGly 948  
  
QY 1078 ATACGGGCCCCAACACACAGCCAAACAAATGAGAGAGTGTGTTGGGACAGCTGGATTTCG 1137  
Db GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys 968  
  
QY 1138 AGACTGTGATCCCGGGGTTTGCAGCTTACAGCCCGCCACTCTACCCGCTAACCAAGAA 1197  
Db ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr 988  
  
QY 1198 AAAGGGGAATTCCTCTGGGCTCTGAGCACACAGAGGCAATTTGATGCTATCAAAAGGCC 1257  
Db GlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTyrGlnGluIleLysGlnAla 1008  
  
QY 1258 CTGCTGAGCGCACCTGCTCTGCGCCCTCCCTCAGCTAACTAAACCTTTTACCTTTATCTG 1317  
Db LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal 1028  
  
QY 1318 GATGAGCGTAAAGGAGTAGCCCGGGGAGTTTAAACCCAAACCTTAGGACCATGAGAAAGA 1377  
Db AspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg 1048  
  
QY 1378 CCTGCTCGCTACCTGTCACAGAACTCGATCCTGTAGCCAGTGGTGGCCCATATGCGCTG 1437  
Db ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTrpProCysLeu 1068  
  
QY 1438 AAGGCTATCGCAGCTGTGGCCATCTACTGCTCAAGGACGCTGACAAATTTGACTTTGGGACAG 1497

```

Db 1069 ArgMetValAlaAlaIleAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGln 1088
QY 1498 AATATAACCTAGTAAAGCCCCCATGCAATGGAGAACATCGTTGGCAGCCCCCAGACCGA 1557
Db 1089 ProLeuValIleLeuAlaProHisAlaValGluAlaLeuValLysGlnProProAspArg 1108
QY 1558 TGGATGACCAACGCCCGCATGACCCCACTATCAAGCCCTGCTTC- --ACAGAGAGGGTC 1614
Db 1109 TrpLeuSerAsnAlaArgMetThrHisTyrGlnAlaLeuLeuLeuLeuAspThrAspArgVal 1128
QY 1615 ACGTTGCTCCACACGCGCTCTCAACCTCCGCTCTTCCTGCTGCTGAAGACACTGTAGAA 1674
Db 1129 GlnPheGlyProValValAlaLeuAsnProAlaThrLeuLeuPro- --LeuProGluGlu 1147
QY 1675 CCAGTGACTCATGATTGGCCATCAACTATTGTATTGAGGAGACTGGGGTCCGCAAGACCTT 1734
Db 1148 GlyLeuGlnHisAsnCysLeuAspIleLeuAlaGluAlaHisGlyThrArgProAspLeu 1167
QY 1735 ACAGACATACCGCTGACTGGAGAGTGTAACTGGTTCACTGACGGAAGCAGTATGTG 1794
Db 1168 ThrAspGlnProLeuProAspAlaAspHisThrTrpTyrThrAspGlySerSerLeuLeu 1187
QY 1795 GTGGAAGCTAAGAGATGGCTGGGCGCGCTGGTGGAGCGGACCCGACCATCTGGGCC 1854
Db 1188 GlnGluGlyGlnArgLysAlaGlyAlaAlaValThrThrGluThrGluValIleTrpAla 1207
QY 1855 AGCAGCTCGCGGAAGGAATCTTACGACCAAAAGCTGAGCTCATGGCCCTCACCAAGCT 1914
Db 1208 LysAlaLeuProAlaGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAla 1227
QY 1915 TTGCGCTGGCGAAGGAATTCATAACATTTATACGACAGCAGCATGCTTTGCG 1974
Db 1228 LeuLysMetAlaGluGlyLysLysLeuAsnValTyrThrAspSerArgTyrAlaPheAla 1247
QY 1975 ACTGCACAGTACATGGGCCATCTATAACAAGGGGTGCTTACTTACCTAGCAGGAGG 2034
Db 1248 ThrAlaHisIleHisGlyGluIleTyrArgArgArgGlyLeuLeuThrSerGluGlyLys 1267
QY 2035 GAAATAAGAACAAAGAGGAATCTTAAGCCTATTAGAAGCGGTACATTTACCAAAAGG 2094
Db 1268 GluIleLysAsnLysAspGluIleLeuAlaLeuLysAlaLeuPheLeuProLysArg 1287
QY 2095 CTAGCTATTATACATGCTCTGTCGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAC 2154
Db 1288 LeuSerIleIleHisCysProGlyHisGlnLysGlyHisSerAlaGluAlaArgGlyAsn 1307
QY 2155 CAGATGCTGACCGGGTGGCCAGCAGCGCCAGCGGTGTAACCTTCTGCTTATATA 2214
Db 1308 ArgMetAlaAspGlnAlaAlaArgLysAlaAla- -----IleThr 1320
QY 2215 GAAATGCCCAAGGCC- -----CCAGAACCCAGA 2241
Db 1321 GluThrProAspThrSerThrLeuLeuIleGluAsnSerSerProTyrThrSerGluHis 1340
QY 2242 CGACAGTACACCTAGAGAGCTGGCAGAGAGATAAAGATAGACACAG- --TTCTCTGAG 2298
Db 1341 PheHisTyrThrValThrAspIleLysAspLeuThrLysLeuGlyAlaIleTyrAspLys 1360
QY 2299 ACTCCGGAAGGGACCTGCTATACCTACATGAGGGGAAGAAATCTGCCCCCACAAGAGGG 2358
Db 1361 ThrLysLysTyrTrpValTyr- -----GlnGlyLysProValMetProAspGlnPheThr 1378
QY 2359 TTAGATATGTCCAAACATACATCGTCTAACCCACCTAGGAACCTAACACCTGCACAG 2418
Db 1379 PheGluLeuLeuAspPheLeuHisGlnLeuThrHisLeuSerPheSerLysMetLysAla 1398
QY 2419 TTGCTCAGAAACA- -----TCCCTTATCATGTTCTGAGGCTTACCAGGAGTGGCTGACTCG 2472
Db 1399 LeuLeuGluArgSerHisSerProTyrTyrMetLeuAsnArgAspArgThrLeuLysAsn 1418
QY 2473 GTGGTCAACATTTGTGGCCCTGCCAGCTGGTTAATGCTAATCTTCCAGATGCGCTCCA 2532
::: ::: |||||:::||||::| ||||| ||||| |||||:::|||||

```

```

Db 1419 IleThrGluThrCysLysAlaCysAlaGlnValAsnAlaSerLysSerAlaValLysGln 1438
QY 2533 GGGAAAGAGACTAAGCGGAGAGCCACCCAGGCGCTCACTGGGAAGTGGACTTCACTCAGAGTA 2592
Db 1439 GlyThrArgValArgGlyHisArgProGlyThrHisTrpGluIleAspPheThrGluIle 1458
QY 2593 AAGCCGCTAAATACGGAACAAATACCTATTGGTTTTTGTGAGACACCTTTTCAGAGTGG 2652
Db 1459 LysProGlyLeuTyrGlyTyrLysTyrLeuLeuValPheIleAspThrPheSerGlyTrp 1478
QY 2653 GTAGAGCTTATCTTACTAAGAAAGACATTCACCCGTGGTGGTCAAAAAATACTGGAA 2712
Db 1479 IleGluAlaPheProThrLysLysGluThrAlaLysValValThrLysLysLeuGlu 1498
QY 2713 GAAATTTTCCAGATTGGAAATACCTAAGGTAAATAGGTGAGCAATGGTCCAGCTTTT 2772
Db 1499 GluIlePheProArgPheGlyMetProGlnValLeuGlyThrAspAsnGlyProAlaPhe 1518
QY 2773 GTGCCAGGTAACTCAGGACTGGCCAAAGATATTGGGGATTGATTGGAATCTCATTTGT 2832
Db 1519 ValSerLysValSerGlnThrValAlaAspLeuLeuGlyIleAspTrpLysLeuHisCys 1538
QY 2833 GCATACAGACCCCAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCATTAAGAGACC 2892
Db 1539 AlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThr 1558
QY 2893 CTTACTAAATTTGACCGGAGACTGGCGTTAATGATTGGATAGCTCTCTGCCCTTTGTG 2952
Db 1559 LeuThrLysLeuThrLeuAlaThrGlySerArgAspTrpValLeuLeuLeuProLeuAla 1578
QY 2953 CTTTATAGGTTAGGAACACCCCTGGACAGTTTGGGCTGACCCCTATGAATTAATCTTAC 3012
Db 1579 LeuTyrArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGluIleLeuTyr 1598
QY 3013 GGGGGACCCCGCCCATTTGGTAGAAATGCTTCTGTACATAGTGTCTGCTGCTCTTCC 3072
Db 1599 GlyAlaProProProLeuValAsnPheProAspProAspMetThrArgValThrAsnSer 1618
QY 3073 CAGCTTTGTTCTTAGGCTCAAGGCACTTGAGTGGGTGAGACACGAGCGTGGAGCAA 3132
Db 1619 ProSerLeuGlnAlaHisLeuGlnAlaLeuTyrLeuValGlnHisGluValTrpArgPro 1638
QY 3133 CTCGGAGAGCCCTACTCAGGAGGAGGACTTGCAG- --ATCCACATCGTTTCCAAAGTG 3189
Db 1639 LeuAlaAlaTyrGlnGluGlnLeuAspArgProValValProHisProTyrArgVal 1658
QY 3190 GGAGATTCACTTACGTTAGACCCCGCTGCGAGGAACCTCGAGACTCGGTGGAAGGGC 3249
Db 1659 GlyAspThrValTrpValArgHisGlnThrLysAsnLeuGluProArgTrpLysGly 1678
QY 3250 CCTTATCTCTACTTTTACACACACCGCTGTGAAGTTCGAAGGAATCTCCACCTGG 3309
Db 1679 ProTyrThrValLeuLeuThrProThrAlaLeuLysValAspGlyIleAlaAlaTrp 1698
QY 3310 ATCCATGCACTCCACGTTAAACCGCGCCACCT- -----CCCGATTCC- --GGG 3354
Db 1699 IleHisAlaAlaHisValLysAlaAlaAspProGlyGlyGlyProSerSerArgLeuThr 1718
QY 3355 TGGAAAGCCGAAAGACTGAAATCCCTTAAGCTTCCCTCCATCGCTGGTTCCT 3411
Db 1719 TrpArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGluAlaPro 1737

```

## RESULT 11

AAV17947

ID AAV17947 standard; protein; 1224 AA.

XX

AC AAV17947;

XX 03-AUG-1999 (first entry)

XX MoMLV pol gene product.

XX

KW Viral packaging signal; chimeric; type C retrovirus; gag gene; MoMLV;

KW murine VL30; retroviral vector; biotechnology; pharmaceutical;  
 KW gene therapy; Moloney murine leukemia virus; pol gene.  
 XX Moloney murine leukemia virus.

XX WO9925862-A2.  
 XX 27-MAY-1999.  
 XX 19-NOV-1998; 98WO-US024667.  
 XX 19-NOV-1997; 97US-0066148P.  
 XX (NATU-) NATURE TECHNOLOGY INC.  
 XX Hodgson CP, Zink MA, Guoping X;  
 DR WPI; 1999-347491/29.  
 DR N-PSDB; AAX77178.  
 XX New chimeric packaging signal useful in gene therapy.  
 PT Disclosure; Fig 4; 6lpp; English.  
 XX The invention relates to a new chimeric viral packaging signal that  
 CC comprises: (i) essential packaging nucleic acid sequence, from a  
 CC mammalian type C retrovirus, functionally joined to (ii). (ii) at least  
 CC one non-essential packaging nucleic acid sequence (lacking the gag gene)  
 CC from murine VL30. The chimeric viral packaging signals are used to  
 CC produce retroviral vectors for transfer of genetic material in gene or  
 CC cell therapy, or other biotechnological applications. Typical uses are in  
 CC production of pharmaceuticals (e.g. enzymes, antibiotics, antibodies,  
 CC antisense RNA, cytokines etc.) in animals or cell cultures. Elimination  
 CC of the gag gene results in safer vectors (reduced formation of  
 CC replication competent retrovirus as a result of recombination events  
 CC involving the gag gene). Also vectors that contain the chimeric packaging  
 CC signal have high infectious titers (about 1 million transduction units  
 CC (TU)/ml), comparable with that for gag-containing vectors and higher than  
 CC for known gag-free vectors. The present sequence represents the pol gene  
 CC product of the MoLV virus  
 XX Sequence 1224 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 1224  
 Score: 3781.00 Matches: 717  
 Percent Similarity: 74.5% Conservative: 165  
 Best Local Similarity: 60.6% Mismatches: 244  
 Query Match: 61.2% Indels: 58  
 DB: 2 Gaps: 11

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x AAY17947 (1-1224)

QY 4 GGTGCCACAGGCAACACAGTATCCATGACCTACCGAAGAACAGTTGACCTGGAGTG 63  
 Db 51 GlyAlaThrGlyGlyLysArgTyrArgTrpThrAspArgLysValHisLeuAlaThr 70  
 QY 64 GCACGGGTAAACCACTCGTTTCTGTCTCATCTGAGTGCACCGACACCCCTCTTAGGTAGA 123  
 Db 71 GlyLysValThrHisSerPheLeuHisValProAspCysProTyrProLeuLeuGlyArg 90  
 QY 124 GACTTATTGACCAAGATGGAGCAAAATTTCTTTTGA---CAAGGGAAACAGAACTG 180  
 Db 91 AspLeuLeuThrLysLeuLysAlaGlnHisPheGluGlySerGlyAlaGlnValMet 110  
 QY 181 TCTGCAATAACAAACCTATCACTGTGTGACCTCCCAATTAGATGACGAATATCGACTA 240  
 Db 111 GlyProMetGlyGlnProLeuGlnValLeuThrLeuAsnIleGluAspGluHisArgLeu 130  
 QY 241 TACTCTCCCTAGTAAAGCCTGATCAAAATATA---CAATTCTGTTTGGAAACAGTTTCCC 297  
 Db 131 HisGluThrSerLysGluProAspValSerLeuGlySerThrTrpLeuSerAspPhePro 150

QY 298 CAAGCTGGGCGAGAAACCGCAGGATGGGTTTGGCAAGCAAGTTCCCCCAAGATTATT 357  
 Db 151 GlnAlaTrpAlaGluThrGlyGlyMetGlyLeuAlaValArgGlnAlaProLeuIleile 170  
 QY 358 CAACTGAAGCCAGTCCACACAGTGTCTAGTCAGACAGTACCCCTTGAGTAAAGAGCT 417  
 Db 171 ProLeuLysAlaThrSerThrProValSerIleLysGlnTyrProMetSerGlnGluAla 190  
 QY 418 CAAGAAGGATTCGGCGCATGTCCAAGATTATCCACAGGCGCATCTAGTTCCCTGTC 477  
 Db 191 ArgLeuGlyIleLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys 210  
 QY 478 CAATCTCCCTGGAATACTCCCTCTACCGGTTAGAAAGCCTGGGACTTAATGACTATCGA 537  
 Db 211 GlnSerProTrpAsnThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg 230  
 QY 538 CCAGTACAGGACTTGAGAGAGGTCAATAACGGGTGAGGATATACACCCCAAGTCCCG 597  
 Db 231 ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro 250  
 QY 598 AACCTTTATACCTCTTGTGTGTCTCCACCCCAACGGAGCTGGTATACAGTATTGGAC 657  
 Db 251 AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp 270  
 QY 658 TTAAGGATGCTTCTTCTGCTGAGATTACACCCCACTAGCCCAACCACTTTTGCCTTC 717  
 Db 271 LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPhe 290  
 QY 718 GAATGGAGAGATCCAGGTACGGGAAGAACCGGGCAGTCTACCTGACCCGACTGCCCAA 777  
 Db 291 GluTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln 310  
 QY 778 GGGTTCAGAACTCCCGACCATCTTTCAGAAAGCCCTACACAGAGACCTGGCCCACTTC 837  
 Db 311 GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe 330  
 QY 838 AGGATCCAAACACCTCAGGTGACCCCTCTCCAGTACGTGGATGACCTGCTTCTGGCGGGA 897  
 Db 331 ArgIleGlnHisProAspLeuIleLeuLeuGlnTyrValAspAspLeuLeuAlaAla 350  
 QY 898 GCCACCAACAGGACTCTTAGAAGGCAACGAGGACCTACTGTGGAAATTTCTGACCTA 957  
 Db 351 ThrSerGluLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAsnLeu 370  
 QY 958 GGCTACAGAGCTCTGCTTAGAAGCCAGATTTCAGAGAGAGAGGTAACTACTTGGGG 1017  
 Db 371 GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly 390  
 QY 1018 TACAGTTTGGCGGACGGGACGATGGCTGACGGAGGCACCGAAGAACTGTAGTCCAG 1077  
 Db 391 TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGluThrValMetGly 410  
 QY 1078 ATACCGGCCCCAACCCACAGCCCAACAAATGAGAGAGATTTTGGGACAGCTGGATTGGC 1137  
 Db 411 GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys 430  
 QY 1138 AGACTGTGGATCCCGGGTTTGGACCTTAGCACCCTTACCCCTTACCCCTAACCAAGAA 1197  
 Db 431 ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr 450  
 QY 1198 AAAGGGGAATTCCTCTGGGCTCTGAGCACCAGGAGCATTTGATGTATCAAAAGGCC 1257  
 Db 451 GlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTyrGlnGluIleLysGlnAla 470  
 QY 1258 CTGTGAGCGCAGCTGTCTTGGCCCTCCTGACGTAACTAAACCTTTTACCTTTATGTG 1317  
 Db 471 LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal 490  
 QY 1318 GATGAGCGTAGGAGTAGCCCGGGGATTTTAAACCCCAACCCCTAGGACCATGAGAGA 1377  
 Db 491 AspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg 510  
 QY 1378 CCTGTGCGCTACCTGTCAAAGAAGCTCGATCCTCTAGCCAGTGGTGGCCCATATGCTG 1437









Db 1740 HisAlaSerHisValLysAlaAlaGlyProThrThrAsnGlnAspLeuSerAspSerPro 1759  
QY 3343 -----CCCGATTCCGGGTGGGAAGCCGAAAGACTGAAATCCCTTAAAGCTTCGC 3393  
Db 1760 SerSerAspAspProSerArgTrpLysValGlnArgThrGlnAsnProLeuLysIleArg 1779  
QY 3394 CTCATCGC 3402  
Db 1780 LeuSerArg 1782  
RESULT 13  
ID AAR05898 standard; protein; 1784 AA.  
XX AAR05898;  
AC AAR05898;  
XX 24-OCT-2003 (revised)  
DT 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 29-NOV-1990 (first entry)  
XX Gene product of first ORF of virus capable of inducing viraemia in feline  
DE species encoded by clone 61E.  
XX Viraemia; leukaemia; FeLV-A; clone 61C; clone 61E; ds.  
XX Feline leukemia virus; strain A.  
OS EP377842-A.  
PN 18-JUL-1990.  
XX 12-DEC-1989; 89EP-0012964.  
XX 13-DEC-1988; 88US-00284139.  
PA (HARD ) UNIV HARVARD.  
PA (COLS ) COLORADO STATE UNIV RES.  
PA (HARD ) UNIV HARVARD.  
XX Hoover EA, Mullins JI;  
XX WPI; 1990-218326/29.  
DR N-PSDB; AAQ05252.  
XX DNA encoding leukaemia virus-A sub: type - producing AIDS type disease in  
PT cats, used to test drugs and vaccines.  
XX Claim 6; Fig 1.1-1.7; 23pp; English.  
XX Variant of viral genome induces immunodeficiency in cats similar to AIDS.  
CC May be used in research, especially in testing drugs and vaccines against  
CC viraemia and feline leukaemia viruses. (Updated on 25-MAR-2003 to  
CC correct PA field.) (Updated on 27-AUG-2003 to correct OS field.) (Updated  
CC on 24-OCT-2003 to standardise OS field)  
XX Sequence 1784 AA;  
SQ  
Alignment Scores:  
Pred. No.: 0 Length: 1784  
Score: 3686.00 Matches: 705  
Percent Similarity: 74.2% Conservative: 158  
Best Local Similarity: 60.6% Mismatches: 258  
Query Match: 59.6% Indels: 42  
DB: 2 Gaps: 12  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x AAR05898 (1-1784)  
QY 4 GGTGCCAGGGCAACACAGTATCCATGGACTACCCGAGAACAGTTGACTTGGAGTG 63  
Db 632 GlyAlaThrGlySerLysAsnTyrArgTrpThrThrAspArgArgValGlnLeuAlaThr 651  
QY 64 GGACGGGTAACCCACTCGCTTCTGTGCATACCTGTAGTGCACGACCCCTCTTAGGTAGA 123

Db 652 GlyLysValThrHisSerPheLeuTyrValProGluCysProTyrProLeuLeuGlyArg 671  
QY 124 GACTTATGACCAAGATGGGAGACCAATTTCTTTTCAA---CAAGGAAACCAAGAGTG 180  
Db 672 AspLeuLeuThrLysLeuLysAlaGlnIleHisPheThrGlyGlyAlaAsnValVal 691  
QY 181 TCTGCAAAATAACAAACCTATCAGCTGTGTGACCTCCAAATTAGATGACGAATATCGACTA 240  
Db 692 GlyProArgGlyLeuProLeuGlnValLeuThrLeuGlnLeuGluGluLysArgLeu 711  
QY 241 TACTCTCCCTAGTAAAGCCCTGATCAAAATATACAATTCTGGTTGGACAGTTTCCCCAA 300  
Db 712 PheGluProGluSerThrGlnLysGlnLysMetAspIleTrpLeuLysAsnPheProGln 731  
QY 301 GCCTGGGCAGAAACCGCAGGATGGGTTTGGCAAGCAAGTTCCCCCAAGTATTTCAA 360  
Db 732 AlaTrpAlaGluThrGlyGlyMetGlyMetAlaHisCysGlnAlaProValLeuIleGln 751  
QY 361 CTGAGGCCAGTGCCACACAGTGTGAGTCAGTCAGACAGTACCCTTGAGTAAAGAGCTCAA 420  
Db 752 LeuLysAlaThrAlaThrProIleSerIleArgGlnTyrProMetProHisGluAlaTyr 771  
QY 421 GAAGGAATTCCGGCCGATGTCCAAAGATTAAATCAACAGGGCATCTCTAGTTCTGTCCAA 480  
Db 772 GlnGlyIleLysProHisIleArgArgMetLeuAspGlnGlyLeuLysProCysGln 791  
QY 481 TCTCCCTGGAATACTCCCTGCTACCGGTTAGAAAGCTGGGACTAATGACTATCGACCA 540  
Db 792 SerProTrpAsnThrProLeuLeuProValLysLysProGlyThrLysAspTyrArgPro 811  
QY 541 GTACAGGACTGTGAGAGGTCAATAAACGGGTGAGGATATACACCAACAGTCCCGAAC 600  
Db 812 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 831  
QY 601 CCTTATAACCTCTTGTGTGCTCTCCACCCCAACCGAGCTGGTATACAGTATTGGACTTA 660  
Db 832 ProTyrAsnLeuLeuSerThrLeuProProSerHisProTyrTyrThrValLeuAspLeu 851  
QY 661 AAGGATGCTTCTTCTGCTGAGATTACACCCCACTAGCCACCACTTTTGCCTTCGAA 720  
Db 852 LysAspAlaPhePheCysLeuArgLeuHisSerGluSerGlnLeuLeuPheAlaPheGlu 871  
QY 721 TGGAGAGATCCAGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTCCCGCAAGGG 780  
Db 872 TrpArgAspProGluIleGlyLeuSerGlyGlnLeuThrTrpThrArgLeuProGlnGly 891  
QY 781 TTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGG 840  
Db 892 PheLysAsnSerProThrLeuPheAspGluAlaLeuHisSerAspLeuAlaAspPheArg 911  
QY 841 ATCCAAACACCTCAGGTGACCTCTCTCAGTACGTGATGACCTGCTTCTGGCGGGAGCC 900  
Db 912 ValArgTyrProAlaLeuValLeuGlnTyrValAspAspLeuLeuLeuAlaAla 931  
QY 901 ACCAAACAGGACTGCTTAGAGGCACGAAGCAGTCTGCTGGAATTGCTCGACCTAGGC 960  
Db 932 ThrArgThrGluCysLeuGluGlyThrLysAlaLeuLeuGluThrLeuGlyAsnLysGly 951  
QY 961 TACAGAGCTCTGTGAAGAGCCAGATTGTGAGGAGAGAGGTAACATACCTTGGGGTAC 1020  
Db 952 TyrArgAlaSerAlaLysLysAlaGlnIleCysLeuGlnGluValThrTyrLeuGlyTyr 971  
QY 1021 AGTTTGGGGACGGGACGAGCTGCTGACGAGGACGCGGAAGAAACTGTAGTCCAGATA 1080  
Db 972 SerLeuLysAspGlyGlnArgTrpLeuThrLysAlaArgLysGluAlaIleLeuSerIle 991  
QY 1081 CCGCCCCCAACCCACAGCCAAACAATGAGAGAGTTTGGGGACAGCTGGGATTTGACAGA 1140  
Db 992 ProValProLysAsnProArgGlnValArgGluPheLeuGlyThrAlaGlyTyrCysArg 1011  
QY 1141 CTGTGGATCCCGGGGTTTGGCAGCCCTTAGCAGCCCTTACCCGCTAACCAAGAAAA 1200

Db 1012 LeuTrpIleProGlyPheAlaGluLeuAlaLaProLeuTyProLeuThrArgProGly 1031  
Qy 1201 GGGGAATTCTCTGGCTCTGAGCAGCAGAGGCAATTGATGTCTATCAAAAGGCCCTG 1260  
Db 1032 ThrLeuPheGlnTrpGlyThrGluGlnGlnLeuAlaPheGluAenIleArgLysAlaLeu 1051  
Qy 1261 CTGAGCGCACTGCTCTGGCCCTCGCTGACGTAACTAAACCTTTACCTTTATGTGGAT 1320  
Db 1052 LeuSerSerProAlaLeuGlyLeuProAspIleThrLysProPheGluLeuPheIleAsp 1071  
Qy 1321 GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCCAAACCTTAGGACCATGAGAGACCT 1380  
Db 1072 GluAsnSerGlyPheAlaLysGlyValLeuValGlnLysLeuGlyProTrpLysArgPro 1091  
Qy 1381 GTCGCCTACCTGTCAAGAAGCTCGATCCTGTAGCCAGTGTGTGGCCCATATGCTGAAG 1440  
Db 1092 ValAlaTyLeuSerLysLysLeuAspThrValAlaSerGlyTrpProProCysLeuArg 1111  
Qy 1441 GCTATCGAGCTGTGGCCATCTGCTGAGCAGCTGACAAATTTGACTTTGGGACAGAA 1500  
Db 1112 MetValAlaAlaIleAlaIleLeuValLysAspAlaGlyLysLeuThrLeuGlyGlnPro 1131  
Qy 1501 ATAACGTAAATAGCCCCCATGCTTGGAGACATCGTTGGCAGCCCCCAGACCGATGG 1560  
Db 1132 LeuThrIleLeuThrSerHisProValGluAlaLeuValArgGlnProProAsnLysTrp 1151  
Qy 1561 ATGACCAACGCCCGCATGACCACTATCAAGACCTGCTTCTC---ACAGAGAGGCTCAG 1617  
Db 1152 LeuSerAsnAlaArgMetThrHisTyGlnAlaMetLeuLeuAspAlaGluArgValHis 1171  
Qy 1617 TTCCTCCACGCGCTCTCAACCTGCGCACTCTTCTGCTGAAGAGACTGATGAACCA 1677  
Db 1172 PheGlyProThrValSerLeuAsnProAlaThrLeuLeuProLeuProSerGlyLysPro 1191  
Qy 1678 -----GTGACTCATGATGCCATCACTATTGATTGAGGAGACTGGGGTCCGCAAGGAC 1731  
Db 1192 ProArgLeuSerProAsp-----LeuAlaGluThrMetAlaGlnThrAsp 1206  
Qy 1732 CTACAGACATACCGCTGACTGGAGAAAGTGTAACTCTGGTTTCACTGACGGAAGCAGCTAT 1791  
Db 1207 LeuThrAspGlnProLeuProAspAlaAspLeuThrTrpTyThrAspGlySerSerPhe 1226  
Qy 1792 GTGGTGAAGATGAAGATGGCTGGGGCGCGGTGGTGGAGCGGACCGCACCATCTGG 1851  
Db 1227 IleArgAsnGlyGluArgLysAlaGlyAlaAlaValThrThrGluSerGluValIleTrp 1246  
Qy 1852 GCAGCAGCTGCGCGAAGGAACCTTACGACAAAGGCTGACTGCTGGCCCTCACGCAA 1911  
Db 1247 AlaAlaSerLeuProProGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGln 1266  
Qy 1912 GCTTTGGCGTGGCGGAAGGMAATCCATAAATTTATACGGACAGCAGTATGCCTTT 1971  
Db 1267 AlaLeuLysMetAlaLysGlyLysLeuThrValTyThrAspSerArgTyAlaPhe 1286  
Qy 1972 GCGACTGCACATGATGGGCGCATCTATAACAAAGGGGGTGTCTTACTCAGCAGGG 2031  
Db 1287 AlaThrAlaHisValHisGlyGluIleTyArgArgGlyLeuLeuThrSerGluGly 1306  
Qy 2032 AGGGAATAAGAACAAAGAGAAATTTAAAGCCTATTAGACCGGTACATTTACCAAAA 2091  
Db 1307 LysGluIleLysAsnLysAsnGluIleLeuAlaLeuLeuGluAlaLeuPheLeuProLys 1326  
Qy 2092 AGGTAGCTATTATACATGCTCTGGACATCAGAAAGCTAATCATCTCATATCCAGAGGA 2151  
Db 1327 ArgLeuSerIleIleHisCysProGlyHisGlnLysGlyAspSerProGlnAlaLysGly 1346  
Qy 2152 AACCATGCTGACCGGGTGGCCAGCAGGACGCC-----CAGGGTGTTPAAC 2199  
Db 1347 AsnArgLeuAlaAspAspThrAlaLysLysAlaAlaThrGluThrGlnSerSerLeuThr 1366  
Qy 2200 CTTCTGCTCT-----ATAATAGAAATGCCCAAGCCCCCAGAACCCAGACAGATACACC 2253  
Db 1367 IleLeuProThrGluLeuIleGluGlyProLys-----ArgProProTrpGluTyArg 1384

Qy 2254 CTAGAGACTGGCAAGAGATAAAAGATAGACACGAGTTCTCTGAGACTCCGGAAGGGACC 2313  
Db 1385 AspSerAspLeuValGlnLysLeuGluAlaHisTyArgGluProLysArgGlyThr 1404  
Qy 2314 TGCTATACCTCAGATGGGAAGAAATCTCTGCCCCACAAAGAGGGTTAGAAATATGTCOA 2373  
Db 1405 ---TrpGluTyArgGlyLysThrIleMetProGluLysTyAlaLysGluLeuIleSer 1423  
Qy 2374 CAGATACATCGTCTAACCCACCTAGGAACATAACACCTGCAGCAGTGGTGGCAGAACATCC 2433  
Db 1424 HisLeuHisLysLeuThrHisLeuSerAlaArgLysMetLysThrLeuLeuGluArgGlu 1443  
Qy 2434 -----CCTTATCATGTTCTTGAGGCTACCAGAGTGGCTGACTCGGTG 2475  
Db 1444 GluThrGlyPheTyLeuProAsnArgAspLeuHisLeuArgGlnValThrGluSer--- 1462  
Qy 2476 GTCMAACATTTGTGTGCCCTGCCAGCTGGTTAATGCTAATCTCTCCAGAAATCCCTCCAGGG 2535  
Db 1463 -----CysArgAlaCysAlaGlnIleAsnAlaGlyLysIleLysPheGlyLeuAsp 1479  
Qy 2536 AAGAGACTAAGGGGAAGCCACCCAGCGCTCACTGGGAAGTGGACTTCACCTGAGGTAAAG 2595  
Db 1480 ValArgAlaArgGlyArgArgProGlyThrHisTrpGluValAspPheThrGluIleLys 1499  
Qy 2596 CCGGCTAAATACGGAAACAAATACCTATTGTTTGTAGACACCTTTTTCAGAGTGGGTA 2655  
Db 1500 ProGlyMetTyArgTyLysTyLeuLeuValPheIleAspThrPheSerGlyTrpAla 1519  
Qy 2656 GAGCTTATCTACTAAGAAGAGACTTCAACCGTGTGGCTAAAAATACTCGAAGAA 2715  
Db 1520 GluAlaTyProAlaLysHisGluThrAlaLysValValAlaLysLysLeuLeuGluGlu 1539  
Qy 2716 ATTTTCCAAAGATTGGAATACCTTAAGGTAAATAGGTGACACATGCTCCGCTTTGCTT 2775  
Db 1540 IlePheProArgTyGlyIleProGlnValLeuGlySerAspAsnGlyProAlaPheIle 1559  
Qy 2776 GCCCAGGTAAAGTCAAGGACTGGCCAGATATTGGGGATTGATTTGGAATCTGCTATGTCGA 2835  
Db 1560 SerGlnValSerGlnSerValAlaThrLeuLeuGlyIleAsnTrpLysLeuHisCysAla 1579  
Qy 2836 TACAGACCCCAAGCTCAGGACAGGTAGAGAGATGAATGAACCATTAAGAGACCTT 2895  
Db 1580 TyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgSerIleLysGluThrLeu 1599  
Qy 2896 ACTAAATTGACCGGAGACTGGCGTTAATGATTGGATAGCTCTCTCCGCTTTGCTGCTT 2955  
Db 1600 ThrLysLeuThrLeuGluThrGlySerLysAspTrpValLeuLeuLeuProLeuValLeu 1619  
Qy 2956 TTTAGGTTAGGAACACCCCTCGACAGTTTGGGCTGACCCCTATGTAATTTACTTACCGG 3015  
Db 1620 TyrArgValArgAsnThrProGlyProHisGlyLeuThrProPheGluIleLeuTyGly 1639  
Qy 3016 GGACCCCGCCCATTTGGTAGAAATTTGCTTCTGTATAGTGTGCTGCTGCTTTTCCAG 3075  
Db 1640 AlaProProMetAlaHisPhePheAspThrAspIleSerThrPheAlaThrSerPro 1659  
Qy 3076 CTTTGTCTCTAGGCTCAAGGCATTTGAGTGGTGGAGACACGAGCGTGGAGGCAATC 3135  
Db 1660 ThrMetGlnAlaHisLeuArgAlaLeuGlnLeuValGlnGluIleGlnArgProLeu 1679  
Qy 3136 CGGAGGCGCTTACTCAGGAGGAGGACTTGCAG---ATCCACATCTGTTTCCAGTGGGA 3192  
Db 1680 AlaAlaAlaTyArgGluLysLeuGluThrProValValProHisProPheLysProGly 1699  
Qy 3193 GATTCACTCTAGCTTAGACCCCGTGGCAGGAAACCTCGAGACTCGGTGGAAGGGCCCT 3252  
Db 1700 AspSerValTrpValArgArgHisGlnThrLysAsnLeuGluProArgTrpLysGlyPro 1719  
Qy 3253 TATCTCGTACTTTTACCACCAACCGGCTGTGAAAGTCAAGGAATCTCCACCTGCATC 3312  
Db 1720 HisIleValLeuLeuThrThrProThrAlaLeuLysValAspGlyValAlaAlaTrpIle 1739



Db 341 GlnLysAlaTyrGlnGluIleLysGlnAlaLeuLeuThrAlaProAlaLeuGlyLeuPro 360  
Qy 1288 GACGTAACAAACCTTTACCTTTATGTGGATGAGCGTAAGGAGTAGCCCGGGAGTT 1347  
Db 361 AspLeuThrLysProPheGluLeuPheValAspGluLysGlnGlyTyrAlaLysGlyVal 380  
Qy 1348 TTAACCCAAACCCTAGGACCAATGAGAAAGACTGTGGCTACCTGTCTCAAGAAAGCTCAT 1407  
Db 381 LeuThrGlnLysLeuGlyProTrpArgArgProValAlaTyrLeuSerLysLysLeuAsp 400  
Qy 1408 CCTGTAGCCAGTGGTGGCCCATATGCTGAAGGCTATCGCAGCTGTGGCCATCTGGTC 1467  
Db 401 ProValAlaAlaGlyTrpProCysLeuArgMetValAlaAlaIleAlaValLeuThr 420  
Qy 1468 AAGACCGCTGCAAAATGACATTTGGGACAGAAATATACTGTAATAGCCCCCATGCATTTG 1527  
Db 421 LysAspAlaGlyLysLeuThrMetGlyGlnProLeuValIleLeuAlaProHisAlaVal 440  
Qy 1528 GAGAACATCGTTCGGCGCCCGACCGATGGATGACCAACGCCCGCATGACCCATAT 1587  
Db 441 GluAlaLeuValLysGlnProProAspArgTrpLeuSerAsnAlaArgMetThrHisTyr 460  
Qy 1588 CAAAGCCTGCTTCTC---ACAGAGAGGTCACGTTTCGCTCCACGCGCTCTCAACCT 1644  
Db 461 GlnAlaLeuLeuLeuAspThrAspArgValGlnPheGlyProValAlaLeuAsnPro 480  
Qy 1645 GCCACTCTTCGCTGGAAGACACTGATGAACAGTGAATCATGATGCTCACTCACTATTG 1704  
Db 481 AlaThrLeuLeuPro---LeuProGluGluGlyLeuGlnHisAsnCysLeuAspIleLeu 499  
Qy 1705 ATTGAGAGACTGGGTCCGACGAGACCTTACAGACATACCGCTGACTGGAGAAAGTGCTA 1764  
Db 500 AlaGluAlaHisGlyThrArgProAspLeuThrAspGlnProLeuProAspAlaAspHis 519  
Qy 1765 ACCTGGTTCTACTGACGGAAGCACTATGTGTGAAGTAAGGATGGCTGGCGCGCG 1824  
Db 520 ThrTrpTyrThrAspGlySerLeuLeuGlnGlyGlnArgLysAlaGlyAlaAla 539  
Qy 1825 GTGTGACGCGGACCCGACCATCTGGCGCAGACGCTGCGGAAGAACTTCAGCACAA 1884  
Db 540 ValThrThrGluThrGluValIleTrpAlaLysAlaLeuProAlaGlyThrSerAlaGln 559  
Qy 1885 AAGCTGAGCTATGCGCCCTCACGCAAGCTTTGGCGTGGCGGAGGAAATCCATAAC 1944  
Db 560 ArgAlaGluLeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLysLysLeuAsn 579  
Qy 1945 ATTTATACGACACAGCTATGCTTTGCGACTGCACACGTACATGCGGCGCATATAAA 2004  
Db 580 ValTyrThrAspSerArgTyrAlaPheAlaThrAlaHisIleHisGlyGluIleTyrArg 599  
Qy 2005 CAAAGGGGTGCTTACTCTCAGCAGGAGGAGAAATAAGAACAAAGGAAATCTTAAGC 2064  
Db 600 ArgArgGlyLeuLeuThrSerGluGlyLysGluIleLysAsnLysAspGluIleLeuAla 619  
Qy 2065 CTATTAGACCGGTACATTTACCAAAAGGCTAGCTATTATACACTGCTCTGACATCAG 2124  
Db 620 LeuLeuLysAlaLeuPheLeuProLysArgLeuSerIleIleHisCysProGlyHisGln 639  
Qy 2125 AAAGCTAAAGATCTCATATCCAGGAAACCAAGTGGCTGACCGGGTTGCCAGCAGGCA 2184  
Db 640 LysGlyHisSerAlaGluAlaArgGlyAsnArgMetAlaAspGlnAlaAlaArgLysAla 659  
Qy 2185 GCCCAGGGTGTAACTTCTGCTATATAAGAAATGCCAAAGCC----- 2229  
Db 660 Ala-----IleThrGluThrProAspThrSerThrLeuLeuIle 672  
Qy 2230 -----CCAGAACCCAGACAGTACACCTTAGAAGACTGGCAAG 2271  
Db 673 GluAsnSerSerProTyrThrSerGluHisPheHisTyrThrValThrAspIleLysAsp 692  
Qy 2272 ATAAAGATAGACACG---TTCTCTGAGACTCCGGAAGGACCTGCTATACCTCATGAT 2328  
Db 693 LeuThrLysLeuGlyAlaIleTyrAspLysThrLysLysTyrTrpValTyr-----Gln 710

Qy 2329 GGGAAAGGAAATCCCTGCCCAACAAAGAGGGTTAGAATATGTCCAACAGATACATCGTCTA 2388  
Db 711 GlyLysProValMetProAspGlnPheThrPheGluLeuLeuAspPheLeuHisGlnLeu 730  
Qy 2389 ACCCACTAGGAATAAACACCTGCAGCAGTGTGTGTCAGAACCA-----TCCCTTATCAT 2442  
Db 731 ThrHisLeuSerPheSerLysMetLysAlaLeuLeuGluArgSerHisSerProTyrTyr 750  
Qy 2443 GTTCTGAGGCTACCAGGAGTGGCTGACTCGTGTGTCAACATTGTGTGCCCTGCCAGCTG 2502  
Db 751 MetLeuAsnArgAspArgThrLeuLysAsnIleThrGluThrCysLysAlaCysAlaGln 770  
Qy 2503 GTTAATGCTAATCCCTCCAGAAATGCTCCAGGGAAGAGACTAAAGGGAAGCCACCCAGGC 2562  
Db 771 ValAsnAlaSerLysSerAlaValLysGlnGlyThrArgValaArgGlyHisArgProGly 790  
Qy 2563 GCTCAGTGGGAGTGGACTTCACCTGAGGTAAAGCCGGCTAAATACGGAAACAATATACCTA 2622  
Db 791 ThrHisTrpGluIleAspPheThrGluIleLysProGlyLeuTyrGlyTyrLysTyrLeu 810  
Qy 2623 TTGTTTTTGTAGACACCTTTTCAGGATGGTAGAGCTTATCTCTACTAAGAAAGAGACT 2682  
Db 811 LeuValPheIleAspThrPheSerGlyTrpIleGluAlaPheProThrLysLysGluThr 830  
Qy 2683 TCACCCGTGTGGCTAAAGAAATACTCGAAAGAAATTTTCCAGATTGTGAATACCTAAG 2742  
Db 831 AlaLysValValThrLysLysLeuLeuGluGluIlePheProArgPheGlyMetProGln 850  
Qy 2743 GTAATAGGTCAGACAAATGTCAGCTTTTGTGGCCAGGTAACTAGTACAGGACTGCCCAAG 2802  
Db 851 ValLeuGlyThrAspAsnGlyProAlaPheValSerLysValSerLysGlnThrValAlaAsp 870  
Qy 2803 ATATTGGGATTCATTTGGAACCTGCTATGCTATACAGACCCCAAGCTCAGGACAGTA 2862  
Db 871 LeuLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnVal 890  
Qy 2863 GAGAGATGAATAGAACCATTAAGACACCTTACTAAATTTGACCGCGAGACTGGCGTT 2922  
Db 891 GluArgMetAsnArgThrIleLysGluThrLeuThrLysLeuAlaThrGlySer 910  
Qy 2923 AATGATTGGATAGCTCTCTGCGCTTTGTGCTTTTATAGGTTAGAACACCCCTGGACAG 2982  
Db 911 ArgAspTrpValLeuLeuLeuProLeuAlaLeuTyrArgAlaArgAsnThrProGlyPro 930  
Qy 2983 TTTGGCTGACCCCTATGAAATCTACTACGGGGACCCCGCCCATTTGTTAGAAATGCT 3042  
Db 931 HisGlyLeuThrProTyrGluIleLeuTyrGlyAlaProProProLeuValAsnPhePro 950  
Qy 3043 TCTGTACATAGTCTGACGTGCTGCTTCCAGCCTTTGCTCTAGGCTCAAGGCACTT 3102  
Db 951 AspProAspMetThrArgValThrAsnSerProSerLeuGlnAlaHisLeuGlnAlaLeu 970  
Qy 3103 GAGTGGGTGAGACACAGCGGTGGAGCAACTCGGAGGCGCTACTCTCAGGAGGAGAGAC 3162  
Db 971 TyrLeuValGlnHisGluValTrpArgProLeuAlaAlaAlaTyrGlnGluGlnLeuAsp 990  
Qy 3163 TTGCAG---ATCCCAACATCGTTTCCAAGTGGGAGATTCAGTCTACGTAGCCGCCCGT 3219  
Db 991 ArgProValValProHisProTyrArgValGlyAspThrValTrpValaArgHisGln 1010  
Qy 3220 GCAGGAAACCTCCAGACTCGGTGGAAGGCGCTTATCTCTGCTATTTTGGACACCAACG 3279  
Db 1011 ThrLysAsnLeuGluProArgTrpLysGlyProTyrThrValLeuLeuThrThrProThr 1030  
Qy 3280 GCTGTGAAGTGAAGAAATCTCCACTGGATCCATGATCCACGTTAAACCGCGGCCA 3339  
Db 1031 AlaLeuLysValAspGlyIleAlaAlaTyrIleHisAlaAlaHisValLysAlaAlaAsp 1050  
Qy 3340 CCT-----CCCCATTG-----GGGTGAAAGCCGAAAGACTGAAATCCCTT 3384  
Db 1051 ProGlyGlyProSerArgLeuThrTrpArgValGlnArgSerGlnAsnProLeu 1070

Qy	3385	AAGCTTCGCTCCATCGCGTGGTCTCT	3411	Qy	1	ATGGGTCCACAGGGCAA---CAACAGTATCCATGGACTACCCGAAGAAGACAGTTGACTTG	57
Db	1071	LysileArgLeuThrArgGluAlaPro	1079	Db	50	IleGlyAlaThrGlyAsnTyrProLysSerGluGlyArgLeuValAspLeu	69
XX	RESULT 15			Qy	58	GGAGTGGGACGGGTAAACCACTCGTTTCTGGTCTACCTGAGTGCCTCCAGACACCCTCTTA	117
XX	ABB79876			Db	70	GlyArgGlyLeuValThrHisSerPheLeuValIleProGluCysProAspProLeuLeu	89
XX	ABB79876 standard; protein; 1193 AA.			Qy	118	GGTAGAGACTATTGACCAAGATGGGAGCAAAATTTCTTTTGAA---CAAGGGAACCA	174
XX	AC	ABB79876;		Db	90	GlyArgAspLeuLeuGlnLysLeuArgAlaThrIleSerPheThrGlyGluGlyProPro	109
XX	25-NOV-2002 (first entry)			Qy	175	GAAGTGTCTGCAATAACAACACCTATCACTGTGTGTGACCTCCCAATTAGATGACGAATAT	234
XX	Spleen necrosis virus DNA polymerase.			Db	110	GluIleArgThrGluGlyLys-----LeuLeuValThrAlaProLeuGluGluGlyTyr	127
XX	Retrovirus; vector; SNV; quiescent cell; gene therapy; vaccine;			Qy	235	CGATATACTCTCCCTAGTAAAGCCCTGATCAAAATATA-----CAATTCCTGGTTG	285
XX	DNA polymerase; enzyme.			Db	128	ArgLeuPhe-----LeuGluAlaProIleGlnAsnValThrLeuLeuGluGlnTrpLys	145
XX	Spleen necrosis virus.			Qy	286	GAACAGTTTCCCAAGCCTGGGCAAGAACCGCAGGGATGGTTGGCAAGCAAGTTCCTCC	345
XX	W0200259338-A2.			Db	146	ArgGluIleProLysValTrpAlaGluIleAsnProProGlyLeuAlaSerThrGlnAla	165
XX	01-AUG-2002.			Qy	346	CCACAAGTTATTCAACTGAAGCCAGTCCACACAGTCAGTCAGTCAGACAGTACCCCTTG	405
XX	07-JAN-2002; 2002WO-US000378.			Db	166	ProIleHisValGlnLeuLeuSerThrAlaLeuProValArgValArgGlnTyrProIle	185
XX	06-JAN-2001; 2001US-0260199P.			Qy	406	AGTAAAGAAGCTCAAGAAGGAATTCGCCCGCATGTCCAAAGATTAAATCCACAGGGCATC	465
XX	(CHIL-) CHILDRENS MEDICAL CENT.			Db	186	ThrLeuGluAlaArgSerLeuArgGluThrIleArgLysPheArgAlaAlaGlyIle	205
XX	(HARD) HARVARD COLLEGE.			Qy	466	CTAGTTCTCTGTCNAATCTCCCTGCAATCTCCCTGCTACCGTTAGAAAGCCCTGGGACT	525
XX	Summerford C, Gray JT, Lee J, Mulligan RC;			Db	206	LeuArgProValHisSerProTrpAsnThrProLeuLeuProValArgLysProGlyThr	225
XX	WPI; 2002-666904/71.			Qy	526	AATGACTATGACACAGTACAGGACTTCGAGAGAGTCAATAAACCGGTGCAGATATACAC	585
XX	N-PSDB; ABQ81150.			Db	226	SerGluTyrArgMetValGlnAspLeuArgGluValAsnLysArgValGluThrIleHis	245
XX	Producing retroviral vector particles for infecting quiescent cells,			Qy	586	CCAAAGTCCCGAACCTTATACCTCTGTGTGCTCTCCACCCCAACGAGCTGGTAT	645
XX	useful in gene therapy comprises co-transfecting cells with a DNA			Db	246	ProThrValProAsnProTyrThrLeuLeuSerLeuLeuProProAspArgThrTrpTyr	265
XX	encoding spleen necrosis virus gag-pol, a heterologous envelope protein,			Qy	646	ACAGTATTGACTTAAAGGATGCTCTTCTGCTGAGATTACACCCCTAGCAACCA	705
XX	and a DNA sequence of interest.			Db	266	SerValLeuAspLeuLysAspAlaPhePheCysIleProLeuAlaProLysSerGlnLeu	285
XX	Disclosure; Fig 8D-N; 68pp; English.			Qy	706	CTTTTTCCTTCGAATGGAGAGATCCAGGTACGGGAAGAACCGGCGAGCTCACCTGGACC	765
XX	The present sequence is the protein sequence of the RNA-dependent DNA			Db	286	IlePheAlaPheGluTyrThrAspAlaGluGluGlyGluSerGlyGlnLeuThrTrpThr	305
XX	polymerase, protease and integrase protein encoded by the pol gene of			Qy	766	CGACTGCCCAAGGGTTCAAGAACTCCCGACCAATCTTTTGACGAAGCCCTTACAGAGAC	825
XX	spleen necrosis virus (SNV). The invention relates to the discovery that			Db	306	ArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAspGluAlaLeuAsnArgAsp	325
XX	retroviral vector particles which encode SNV gag-pol gene products can			Qy	826	CTGCCCAACTTCAGATCCACACCTCAGTGACCTCTCCAGTACGTGGATGACCTG	885
XX	infect and transduce a DNA sequence of interest into quiescent (non-			Db	326	LeuGlnGlyPheArgLeuAspHisProSerValSerLeuLeuGlnTyrValAspAspLeu	345
XX	dividing, resting, non-proliferating) cells. Production of a retroviral			Qy	886	CTTCTGGCGGAGCCACCAACAGGACTGCTTAGAAGGCACGAAGGCACTACTGCTGAA	945
XX	vector particle capable of infecting quiescent cells comprises co-			Db	346	LeuIleAlaAlaAspThrGlnAlaAlaCysLeuSerAlaThrArgAspLeuLeuMetThr	365
XX	transfecting mammalian host cells with: (a) a first plasmid containing a			Qy	946	TTGTCTACCTTAGGCTACAGACCTCTGCTTAAGAAGGCCAGATTTCGAGAGAGAGTA	1005
XX	DNA sequence encoding wild type SNV gag-pol; (b) a second plasmid			Db	366	LeuAlaGluLeuGlyTyrArgValSerGlyLysLysAlaGlnLeuCysGlnGluGluVal	385
XX	containing a DNA sequence encoding a heterologous envelope protein; and			Qy	1006	ACATACTGGGGTACAGTTTCGGGACCGGCGAGCGATGGCTGACGCGAGCGGAAGAAA	1065
XX	(c) a third plasmid containing a DNA sequence of interest. Novel			Db	386	ThrTyrLeuGlyPheLysIleHisLysGlySerArgThrLeuSerAsnSerArgThrGln	405
XX	packaging cell lines useful for generating the retroviral vector			Qy	1066	ACTGTAGTCCAGATACCGGCCCCCAACACACCAACAAATAGAGAGATTTTGGGGACA	1125
XX	particles are claimed. A method of gene transfer to quiescent cells using						
XX	these retroviral vector particles is also claimed. The packaging cell						
XX	lines and the viral particles can be used in gene therapy or gene						
XX	replacement to introduce genes into a variety of quiescent cells, in the						
XX	development and production of vaccines, and in the production of						
XX	biochemical agents						
XX	Sequence 1193 AA;						
XX	Alignment Scores:						
XX	Pred. No.:	2,99e-266	Length:	1193			
XX	Score:	3090.00	Matches:	623			
XX	Percent Similarity:	67.3%	Conservative:	158			
XX	Best Local Similarity:	53.7%	Mismatches:	332			
XX	Query Match:	50.0%	Indels:	48			
XX	DB:	5	Gaps:	15			
XX	US-10-723-552-3_COPY_2307_5741 (1-3435) x ABB79876 (1-1193)						



```
Qy 3220 GCAGGAACCTCGAGACTCGGTGAAGGGCCCTTATCTCGTACTTTTACCCAGACCAACG 3279
Db 1129 PheGlnGlnLeuGlyProArgTTPAspGlyProTyrThrValValLeuSerThrProThr 1148
Qy 3280 GCTGTGAAGTCGAAGGAATCTCCACCTGGATCCATGCCACGTTAAACCGCGGCCA 3339
Db 1149 AlaValLysValAlaGlyLysThrProTrpIleHisSerArgLeuLysLysAlaPro 1168
Qy 3340 CCTCCCGATTCCGGGTGGAAGCCGAAAGACTGAAAATCCCTTAAGCTTCGCCTCCAT 3399
Db 1169 AspAsnGlnGluGluTyrThrValSerProThrSerAspProLeuArgValLysLeuThr 1188
Qy 3400 CGC 3402
Db 1189 Arg 1189
```

Search completed: February 14, 2006, 15:54:26  
Job time : 670.958 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 15:55:36 ; Search time 17.7535 Seconds  
(without alignments)  
3199.272 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_2307\_5741

Perfect score: 6183

Sequence: 1 ATGGGTCCACAGGCAACA.....CTGTCAATAACCTCTCAGAC 3435

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB.spool/US10723552/runat 14022006 125146 13079/app.query.fasta.1  
-DB=Issued\_Patents\_AA -QMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs802p -USER=US10723552 @CGN 1.1 161 @runat 14022006 125146 13079  
-NCPU=6 -ICPU=3 -NO MWAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2\_6/ptodata/1/iaa/5 COMB.pap.\*
- 2: /cgn2\_6/ptodata/1/iaa/6 COMB.pap.\*
- 3: /cgn2\_6/ptodata/1/iaa/H COMB.pap.\*
- 4: /cgn2\_6/ptodata/1/iaa/PCUTS COMB.pap.\*
- 5: /cgn2\_6/ptodata/1/iaa/RE COMB.pap.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5851.5	94.6	1194	2	US-09-171-553B-5
2	4143.5	67.0	1203	2	US-09-075-272-4
3	3797.5	61.4	1737	2	US-09-309-572-13
4	3797.5	61.4	1737	2	US-09-718-096-13
5	3600	58.2	1079	1	US-08-929-967-8
6	2500.5	40.4	665	1	US-08-929-967-7
7	1501.5	24.3	379	2	US-09-603-185-6
8	1415.5	22.9	768	2	US-08-979-847B-89
9	1390.5	22.5	683	2	US-08-979-847B-198
10	1390.5	22.5	683	2	US-08-979-847B-208
11	1387.5	22.4	683	2	US-08-979-847B-208
12	1387.5	22.4	683	2	US-08-979-847B-210

13	1331.5	21.5	654	2	US-08-979-847B-91
14	1236	20.0	768	2	US-09-120-653D-5
15	629.5	10.2	917	2	US-08-259-451-11
16	579	9.4	995	4	PCT-US95-04910-14
17	578.5	9.4	1802	2	US-09-322-478-18
18	578.5	9.4	1802	2	US-09-586-106D-18
19	578.5	9.4	1802	2	US-10-799-870-18
20	528	8.5	1150	2	US-09-238-303-9
21	528	8.5	1150	2	US-09-946-239-9
22	512.5	8.3	1055	1	US-08-659-251-5
23	512.5	8.3	1055	2	US-09-256-490-5
24	512.5	8.3	1055	4	PCT-US96-11445-5
25	512	8.3	1122	2	US-09-248-796A-16374
26	508	8.2	1031	2	US-08-811-682-15
27	506.5	8.2	1014	2	US-09-319-588C-6
28	493.5	8.0	858	2	US-09-960-428-7
29	491	7.9	1018	2	US-09-206-551-46
30	487.5	7.9	3080	6	5223423-4
31	473.5	7.7	1016	2	US-09-625-972-24
32	470	7.6	1003	2	US-10-290-579A-193
33	460	7.4	1002	2	US-10-290-579A-189
34	460	7.4	1003	1	US-07-743-357-9
35	460	7.4	1005	2	US-10-290-579A-191
36	458.5	7.4	1003	2	US-10-290-579A-192
37	455.5	7.4	1003	2	US-10-290-579A-183
38	452.5	7.3	913	1	US-07-743-357-6
39	450	7.3	1000	2	US-10-290-579A-186
40	449	7.3	999	2	US-10-290-579A-190
41	446	7.2	1003	1	US-07-743-357-8
42	446	7.2	1004	1	US-07-743-357-7
43	443	7.2	1003	2	US-09-393-795-9
44	443	7.2	1003	2	US-09-393-795-11
45	440	7.1	1003	1	US-07-743-357-10

ALIGNMENTS

RESULTS  
US-09-171-553B-5  
; Sequence 5, Application US/09171553B  
; Patent No. 6756227  
; GENERAL INFORMATION:  
; APPLICANT: GALBRAITH, DANIEL N.  
; APPLICANT: HAWORTH, CHRISTINE  
; APPLICANT: LEES, GILLIAN M.  
; APPLICANT: SMITH, KENNETH T.  
; TITLE OF INVENTION: PORCINE RETROVIRUS  
; FILER REFERENCE: CFV-5.01  
; CURRENT APPLICATION NUMBER: US/09/171.553B  
; CURRENT FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: PCT/GB97/01087  
; PRIOR FILING DATE: 1997-04-18  
; PRIOR APPLICATION NUMBER: GB 9702668.6  
; PRIOR FILING DATE: 1997-02-10  
; PRIOR APPLICATION NUMBER: GB 9608164.1  
; PRIOR FILING DATE: 1996-04-19  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1194  
; TYPE: PRT  
; ORGANISM: Porcine retrovirus  
US-09-171-553B-5

Alignment Scores:  
Pred. No.: 0  
Score: 5851.50  
Percent Similarity: 98.4%  
Best Local Similarity: 96.9%  
Query Match: 94.6%  
DB: 2  
Length: 1194  
Matches: 1107  
Conservative: 17  
Mismatches: 17  
Indels: 1  
Gaps: 1

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-09-171-553B-5 (1-1194)

QY 1 ATGGGTGCCAGGCGCAACACAGTATCCATGGACTACCCGAAACAGTGTGACTTGGGA 60  
Db 50 MetGlyAlaThrGlyGlnArgGlnProThrThrArgThrValAspLeuGly 69  
QY 61 GTGGACCGGGTAACCCACTCGCTTTCTGTGTATACCTGAGTCCCGCAGCACCCCTCTTAGGT 120  
Db 70 ValGlyArgValThrHisSerPheLeuValIleProGluCysProValProLeuLeuGly 89  
QY 121 AGAGACTTATTGACCAAGATGGGAGCAAAATTTCTTTTGAACAGGGAACCAAGAGTG 180  
Db 90 ArgAspLeuLeuThrLysMetGlyAlaGlnIleSerPheGluGlnGlyArgProGluVal 109  
QY 181 TCTCCAATACAAACCTATCACTGTGTGACCTCCAAATTAGATGAGATATCGACTA 240  
Db 110 SerValAsnAsnLysProIleThrValLeuThrLeuGlnLeuAspAspGluTyArgLeu 129  
QY 241 TACTCTCCCTAGTAAGCCTGATCAAAATATACAAATTTCTGTTGGAAACAGTTCCTCCCAA 300  
Db 130 TyrSerProGlnValLysProAspGlnAspIleGlnSerThrLeuGluGlnPheProGln 149  
QY 301 GCCTGGCAGAAACCGCAGGATGGTTTGGCAAGCAAGTTCCCCCAAAAGTTATTCAA 360  
Db 150 AlaTrpAlaGluThrAlaGlyMetGlyLeuAlaLysGlnValProProGlnValIleGln 169  
QY 361 CTGAAGCCAGTCCACACAGTGTCTAGTCAGACAGTACCCCTTGAGTAAGAAGCTCAA 420  
Db 170 LeuLysAlaSerAlaThrProValSerValArgGlnTyrProLeuSerArgGluAlaArg 189  
QY 421 GAAGGAATTCGGCGCATGTCCAAAGATTAAATCAACAGGCATCTAGTTCTCTGTCCAA 480  
Db 190 GluGlyIleTrpProHisValGlnArgLeuIleGlnGlnGlyIleLeuValProValGln 209  
QY 481 TCTCCCTGGAAATCTCCCTGCTACCGGTTAGAAAGCCTGGGACTAATGACTATCGACCA 540  
Db 210 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyArgPro 229  
QY 541 GTACAGGACTTGAAGAGGTCAATAACGGGTGAGGATATACACCAACAGTCCCGAAC 600  
Db 230 ValGlnAspLeuArgGluValAsnLysArgValGlnAspIleHisProThrValProAsn 249  
QY 601 CCTTATAACCTCTTGTGTGCTCTCCACCCCAACGGAGCTGGTATACAGTATTGGACTTA 660  
Db 250 ProTyrAsnLeuLeuSerAlaLeuProProGluArgAsnTrpTyrThrValLeuAspLeu 269  
QY 661 AAGGATGCTTCTTCTGTGCTGAGATTACACCCCACTAGCCAAACCACTTTTGCCTTCGAA 720  
Db 270 LysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGlu 289  
QY 721 TGGAGAGATCCAGGTACGGGAGAACCGGCGAGCTCACCTGGACCCGACTGCCCAAGG 780  
Db 290 TrpArgAspProGlyThrGlyArgThrGlyGlnLeuThrTrpThrArgLeuProGlnGly 309  
QY 781 TTCGAAGAACTCCCGACCATCTTTGACGAAGCCCTTACACAGAGACTTGGCCAACTTCAGG 840  
Db 310 PheLysAsnSerProThrIlePheAspGluAlaLeuHisArgAspLeuAlaAsnPheArg 329  
QY 841 ATCCAACACCTCAGGTACCCCTCTCCAGTACGTGGATGACCTGCTTCTGGCGGAGCC 900  
Db 330 IleGlnHisProGlnValThrLeuLeuGlnTyrValAspAspLeuLeuLeuAlaGlyAla 349  
QY 901 ACCAAACAGGACTGCTTAGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960  
Db 350 ThrLysGlnAspCysLeuGluGlyThrLysAlaLeuLeuLeuGluLeuLeuSerAspLeuGly 369  
QY 961 TACAGAGCTCTGCTAAGAAGCCAGATTTTGCAGGAGAGAGGTAACATACTTGGGGTAC 1020  
Db 370 TyrArgAlaSerAlaLysLysAlaGlnIleCysArgArgGluValThrTyrLeuGlyTyr 389  
QY 1021 AGTTTGGGACGGGACGCGATGGCTGACGAGGACGCGAAGAAACCTGTAGTCCAGATA 1080  
Db 390 SerLeuArgGlyGlyGlnArgTrpLeuThrGluAlaArgLysLysThrValValGlnIle 409

QY 1081 CCGGCCCAACCAACAGCAAAATGAGAGAGTTTTTTGGGGAAGTGTGATTTTGCAGA 1140  
Db 410 ProAlaProThrThrAlaLysGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 429  
QY 1141 CTGTGGATCCCGGGTTTGGACCTTAGCAGCCCACTTACCCGCTAACCAAGAAAAA 1200  
Db 430 LeuTrpIleProGlyPheAlaThrLeuAlaProLeuTyrProLeuThrLysGluLys 449  
QY 1201 GGGGAATTCCTCGGGCTCTGAGCACAGAGCATTTGATGCTATCAAAAGGCCCTG 1260  
Db 450 GlyGlyPheSerTrpAlaProGluHisGlnLysAlaPheAspAlaIleLysLysAlaLeu 469  
QY 1261 CTGAGCCACCTGCTCGGCCCTCCCTGACGTAAACCTTTTACCTTTATGTGGAT 1320  
Db 470 LeuSerAlaProAlaLeuAlaLeuProAspValThrLysProPheThrLeuTyrValAsp 489  
QY 1321 GAGCGTAAGGAGTAGCCCGGGAGTTTAAACCAACCTTAGGACCATGAGAGACCT 1380  
Db 490 GluArgLysGlyValAlaArgGlyValLeuThrGlnThrLeuGlyProTrpArgArgPro 509  
QY 1381 GTCCCTACCTGTCAAGAAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATGCTGAAG 1440  
Db 510 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProValCysLeuLys 529  
QY 1441 GCTATCCAGCTGTGGCCATACTGTCAGAGCGCTGACAAATGACTTTGGGACAGAAT 1500  
Db 530 AlaIleAlaValAlaIleLeuValLysAspAlaAspLysLeuThrLeuGlyGlnAsn 549  
QY 1501 ATAACTGTAATAGCCCCCATGTCATGAGAACATCGTTGGGAGCCCGCAGACCGATGG 1560  
Db 550 IleThrValIleAlaProHisAlaLeuGluAsnIleValArgGlnProProAspArgTrp 569  
QY 1561 ATGACCAACGCCGCATGACCCACTATCAAGCTGCTCTCACAGAGAGGTCACGCTC 1620  
Db 570 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuThrGluArgValThrPhe 589  
QY 1621 GCTCCACAGCCGCTCTCAACCCCTGCCACTCTTCTGCTGAAGAGACTGATGAACCACTG 1680  
Db 590 AlaProProAlaAlaLeuAsnProAlaThrLeuLeuProGluGluThrAspGluProVal 609  
QY 1681 ACTCATGATTGCCATCAACTATTGATTGAGAGACTGGGGTCCGACAGGACCTTACAGAC 1740  
Db 610 ThrHisAspCysHisGlnLeuLeuIleGluIleGluThrGlyValArgLysAspLeuThrAsp 629  
QY 1741 ATACCGTGTACTGAGAGTCTTAACCTGTTCACTGACGGAAGCAGCTAGTGTGTGAA 1800  
Db 630 IleProLeuThrGlyGluValLeuThrTrpPheThrAspGlySerSerTyrValValGlu 649  
QY 1801 GGTAAAGAGATGGCTGGGGCGGCTGTGAGCCGGGACCCCGCAGCATCTGGGCCAGCAGC 1860  
Db 650 GlyLysArgMetAlaGlyAlaAlaValValAspGlyThrArgThrIleTrpAlaSerSer 669  
QY 1861 CTGCCGGAAGAACTTCAGCACAAAGCTGAGTCAATGGCCCTCAGCAAGCTTTGGCG 1920  
Db 670 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuMetAlaLeuThrGlnAlaLeuArg 689  
QY 1921 CTGCCCGAGGGAATCCATAAACATTATACGACAGCAGGTATGCTTTGCCACTGCA 1980  
Db 690 LeuAlaGluGlyLysSerIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 709  
QY 1981 CACGTACATGGGGCATCTATAAAAGGGGGTGTGTACTCTCAGCAGGAGGAGGAATA 2040  
Db 710 HisValHisGlyAlaIleTyrLysGlnArgGlyLeuLeuThrSerAlaGlyArgGluIle 729  
QY 2041 AAGAACAAAGAGGAAATTTCTAAGCCTATTAGAAGCCGTACATTTTACCAAAAGCGTAGCT 2100  
Db 730 LysAsnLysGluGluIleLeuSerLeuLeuGluAlaLeuHisLeuProLysArgLeuAla 749  
QY 2101 ATTATACACTGCTCGGACATCAGAAAGCTTAAGATCTCATATCCAGAGGAAACAGATG 2160  
Db 750 IleIleHisCysProGlyHisGlnLysAlaLysAspLeuIleSerArgGlyAsnGlnMet 769  
QY 2161 GCTGACCGGGTTGCCAAAGCAGGCCAGGGTGTTAACCTTCTGCTCTATTAAGAAATG 2220

Qy	1	ATGGGTGCCACAGGCCAACACAGTATCCATCGGACTACCCGAGAACACGTTGACTTGGGA	60
		::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::	
Db	50	ValGlyAlaThrGlySerLysLysLeuTyrProTrpThrThrLysArgAlaLeuGlnIleAsp	69

QY 61 GTGGAGCGGTAACCCACTCGTTCTGTGTCATCTAGTGGCCAGCACCCCTCTTAGGT 120  
Db |||||  
QY 70 LysAsnMetValThrHisSerPheLeuValIleProGluCysProAlaProLeuLeuGly 89  
Db |||||  
QY 121 AGAGACTTATTGACCAAGATGGGACCAAAATTTCTTTGAACAGGGGAAACCAAGAGTG 180  
Db |||||  
QY 90 ArgAspLeuLeuThrLysLeuLysAlaGlnValGlnPheThrSerGluGlyProGlnVal 109  
Db |||||  
QY 181 TCTGCAATAACAACCTATCACTGTGTGTGACCTCCAAATTAGATGACGATATCGACTA 240  
Db |||||  
QY 110 SerTrpGlyLysAlaProLeuAlaCysLeuValLeuSerThrGluGluGluTyrArgLeu 129  
QY |||||  
QY 241 TACTCTCCCTAGTAAGCCTGATCAAAATATACAAATTCGTGTTGGACAGCTTCCCAA 300  
Db |||||  
QY 130 HisGluGluGlnProLysGlyAlaAlaProLeuAsp---TrpValThrAlaPheProAsn 148  
QY 301 GCCTGGGAGAAAACCGCAGGGATGGTTTGGCAAGCAAGTTCCTCCCAACAAGTTATTCAA 360  
Db |||||  
QY 149 ValTrpAlaGluGlnAlaGlyMetGlyLeuAlaLysGlnValProProValValValGlu 168  
QY 361 CTGAAGCGCAGTGCACACAGTGTCACTGACAGACGATACCCCTTGAGTAAGAAGCTCAA 420  
Db |||||  
QY 169 LeuLysAlaAspAlaThrProIleSerValArgGlnTyrProMetSerLysGluAlaLys 188  
QY 421 GAAGGAATTCGGCCGATGTCCAAAGATTAAATCCAAACAGGCATCCTAGTTCTCTGCCAA 480  
Db |||||  
QY 189 GluGlyIleAspProHisIleArgArgLeuLeuAspGlnGlyIleLeuValAlaCysGln 208  
QY 481 TCTCCTCGAATATCTCCCTGCTACCGTTAGAAAGCCTGGGACTTAATGACTATCGACCA 540  
Db |||||  
QY 209 SerProTrpAsnThrProLeuLeuProValArgLysProGlyThrAsnAspTyrArgPro 228  
QY 541 GTACAGGACTTGAGAGAGTGTAATAACGGGTGCAGGATATACACCCAAACAGTCCCGAAC 600  
Db |||||  
QY 229 ValGlnAspLeuArgGluValAsnLysArgValLeuAspIleHisProThrValProAsn 248  
QY 601 CCTTATACCTCTGTGTGCTCTCCACCCCAACGGAGCTGTATACAGTATTGGACTTA 660  
Db |||||  
QY 249 ProTyrAsnLeuLeuSerSerLeuProGluArgThrTrpTyrThrValLeuAspLeu 268  
QY 661 AAGGATGCTCTTCTTGCTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTTCGAA 720  
Db |||||  
QY 269 LysAspAlaPhePheCysLeuArgLeuHisProLysSerGlnLeuLeuPheAlaPheGlu 288  
QY 721 TGGAGAGATCAGGTACGGGAAGAACCGGGCAGCTCACTGACCCGACTGCCCAAGGG 780  
Db |||||  
QY 289 TrpArgAspProGluGlyGlyGlnThrGlyGlnLeuThrTrpTrpArgLeuProGlnGly 308  
QY 781 TTCAGAACTCCCGACCATCTTTCACCAAGCCCTACACAGAGACTTGCACAACCTTCAGG 840  
Db |||||  
QY 309 PheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaProPheArg 328  
QY 841 ATCCAAACACCTCAGGTGACCCCTCCACGATACGTGTGATGACCTGCTTCTGGCGGAGCC 900  
Db |||||  
QY 329 AlaGlnAsnProGlnLeuThrLeuLeuGlnTyrValAspAspLeuLeuIleAlaAla 348  
QY 901 ACCAAACAGGACTGCTTAGAGGACGAAGGACGACTACTGTGGAATTTGCTGACCTAGGC 960  
Db |||||  
QY 349 SerLysGluLeuCysGlnGlnGlyThrGluArgLeuLeuThrGluGlyAsnLeuGly 368  
QY 961 TACAGAGCTCTGCTAAGAGGCCAGATTTGACGAGAGAGGTTAACTACTTGGGGTAC 1020  
Db |||||  
QY 369 TyrArgValSerAlaLysLysAlaGlnIleCysGlnThrGluValIleTyrLeuGlyTyr 388  
QY 1021 AGTTTGGGGAGGGCAGCGATGGCTGACGGAGGACGGAAGAACTGTAGTCAGATA 1080  
Db |||||  
QY 389 ThrLeuArgGlyGlyLysArgTrpLeuThrGluAlaArgLysLysThrValMetMetIle 408  
QY 1081 CCGGCCCCAACACAGCCAAACAATAGAGAGATTTTGGGACAGCTGGATTTGGCAGA 1140  
Db |||||  
QY 409 ProProProThrThrProArgGlnValArgGluPheLeuGlyThrAlaGlyPheCysArg 428  
QY 1141 CTGTGATCCCGGGTTTGGACCTTAGCAGCCCTTACCGCTCTACCCGCTAACCAAGAAAA 1200

Db |||||  
QY 429 LeuTrpIleProGlyPheAlaThrLeuAlaAlaProLeuTyrProLeuThrArgGluGly 448  
QY 1201 GGGGAATTCCTCGGCTCTGAGCAGCAGAGGCATTTGATGCTATCAAAAAGGCCCTG 1260  
Db |||||  
QY 449 IleProPheGluTrpLysGluGluHisGlnArgAlaPheGluAlaIleLysSerSerLeu 468  
QY 1261 CTGAGCCACCTGCTCGGCCCTCCCTGACGTAACCAACCCCTTTACCTTTATGTGAT 1320  
Db |||||  
QY 469 MetThrAlaProAlaLeuAlaLeuProAspLeuThrLysSerPheValLeuTyrValAsp 488  
QY 1321 GAGCGTAGGAGTAGCCCGGGAGTTTAAACCCCAACCCCTTAGGACCATGAGAGACCT 1380  
Db |||||  
QY 489 GluArgAlaGlyIleAlaArgGlyValLeuThrGlnAlaLeuGlyProTrpLysArgPro 508  
QY 1381 GTCCCTACTGTCAAAAGAGCTCGATCCGTAGCCAGTGGTGGCCCATATGCTGTAAG 1440  
Db |||||  
QY 509 ValAlaTyrLeuSerLysLysLeuAspProValAlaSerGlyTrpProThrCysLeuLys 528  
QY 1441 GCTATCCAGCTGGCCATACTGTCAGAGACCTGACAAATTGACTTTTGGACAGAAT 1500  
Db |||||  
QY 529 AlaIleAlaValAlaLeuLeuIleLysAspAlaAspLysLeuThrMetGlyGlnGln 548  
QY 1501 ATAACTGTAAATAGCCCCCATGTCATTCGAGAACATCGTTTCGGCAGCCCCCAGACCGATGG 1560  
Db |||||  
QY 549 ValThrValValAlaProHisAlaLeuGluSerIleValArgGlnProProAspArgTrp 568  
QY 1561 ATGACCAACGCCCGCATGACCCACTATCAAAAGCCTGCTTTCACAGAGAGGTCACGTTTC 1620  
Db |||||  
QY 569 MetThrAsnAlaArgMetThrHisTyrGlnSerLeuLeuLeuAsnAspArgValThrPhe 588  
QY 1621 GCTCACCGCCGCTCTCAACCCCTGCCACTCTCTGCTGGAAGACTGTATGAACCACTG 1680  
Db |||||  
QY 589 AlaProProAlaIleLeuAsnProAlaThrLeuLeuProLeuThrAsnAspSerValPro 608  
QY 1681 ACTCATGATTCCTCACTTATTGATGAGGAGACTGGGGTCCGCAAGGACCTTACAGAC 1740  
Db |||||  
QY 609 ValHisArgCysAlaAspIleLeuAlaGluGlyIleGlyThrArgLysAspLeuThrAsp 628  
QY 1741 ATACCGTGAAGTGGAGAGTCTAACTGGTTCCTGACGGAAGCAGCTATGTGGTGAA 1800  
Db |||||  
QY 629 GlnProTrpProGly---AlaProSerTrpTyrThrAspGlySerSerPheLeuIleGlu 647  
QY 1801 GGTAAAGAGTAGGTGGGGCGGTGGTGACGGGACCCGACATCTGGCGCAGCAGC 1860  
Db |||||  
QY 648 GlyLysArgArgAlaGlyAlaValValAspGlyLysValIleTrpAlaSerAla 667  
QY 1861 CTGCCGGAAGAACTTCAGCACAAAAGCTCAGCTCATGGCCCTCACCAAGCTTTGCGG 1920  
Db |||||  
QY 668 LeuProGluGlyThrSerAlaGlnLysAlaGluLeuIleAlaLeuThrGlnAlaLeuArg 687  
QY 1921 CTGCCGGAAGGAAATCCATAAACATTTATACGACAGCAGCTATGCTTTGCGACTGCA 1980  
Db |||||  
QY 688 GluAlaGluGlyLysIleIleAsnIleTyrThrAspSerArgTyrAlaPheAlaThrAla 707  
QY 1981 CACGTACATGGGCGCATCTATAAAAGGGGTTGCTTACTCAGCAGGAGGAGGAAATA 2040  
Db |||||  
QY 708 HisIleHisGlyAlaIleTyrArgGlnArgGlyLeuLeuThrSerAlaGlyLysAspIle 727  
QY 2041 AAGAACAAAGAGGAAATTTCTAAGCCTATTAGAAGCCGTACATTTTACCAAAAAGGCTAGCT 2100  
Db |||||  
QY 728 LysAsnLysGluGluIleLeuAlaLeuLeuGluAlaIleHisAlaProLysLysValAla 747  
QY 2101 ATTATACACTGCTCGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCATG 2160  
Db |||||  
QY 748 IleIleHisCysProGlyHisGlnLysGlyGlyLeuValAlaLysGlyAsnArgMet 767  
QY 2161 GCTACCGGGTTGCCAGCAGCCAGCGGTGTAACTTCTGCTCTATATAGAAATG 2220  
Db |||||  
QY 768 AlaAspSerValAlaLysGlnValAlaGlnGlyAlaMetIleLeuThrGluLysGlyAsn 787  
QY 2221 CCC---AAAGCCCAAGAACCCAGCAGCAGTACACCTTAGAAGACTGGCAAGAGATAAAA 2277  
Db |||||

```
Db      788 ProSerLysSerProGlu-----AspGluAsnTyr 797
Qy      2278 AAGATAGACCAAGTTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGGAAGAA 2337
Db      798 AspiLeLysGluLeuPheThrTyrSerAspProLeuProTyrPhePheGluGlyIle 817
Qy      2338 ATCTGCCCCCAAGAAGGGTTAGATATGTCACAGATACATGCTTAACACCACTA 2397
Db      818 AspLeuThrProGluGluGlyIleLysPheValLysGlyLeuHisGlnPheThrHisLeu 837
Qy      2398 GGAATTAACACCTGCGTGCAGACATCCCTTATCATGCTTCTGAGGCTACCA 2457
Db      838 GlyValGluLysMetMetArgLeuIleLysLysSerArgTyrGlnValProAsnLeuLys 857
Qy      2458 GGAGTGGTGAAGTGGTGCCTCAACATTTGTCCTGCCAGCTGGTTAATGCTAATCCT 2517
Db      858 SerValAlaGlnLysIleAsnSerCysLysAlaCysAlaPheThrAsnAlaThrLys 877
Qy      2518 TCCAGATGCTCCAGGGAAGAGACTAAGGGGAAGCCACCAGGCGCTCACTGGGAAGTG 2577
Db      878 ThrTyrLysGluProGlyLysArgGlnArgGlyAspArgProGlyValTyrTrpGluVal 897
Qy      2578 GACTTCTACTAGGTAAAGCGGTAAATACGGAACAAATACCTATTGTTTGTAGAC 2637
Db      898 AspPheThrGluValLysProGlyMetTyrGlyAsnLysTyrLeuLeuValPheValAsp 917
Qy      2638 ACCTTTTTCAGGATGGTAGAGCTTATCTACTAAGAAAGAGACTTCAACCGTGGTGGCT 2697
Db      918 ThrPheSerGlyTrpValGluAlaPheProThrLysThrGluThrAlaGlnIleValAla 937
Qy      2698 AAAAAATACTGGAAGAAATTTTCCAGATTTGGAATACCTAAGTAAGTAAAGGTTCAGAC 2757
Db      938 LysLysIlePheGluGluLeuLeuProArgTyrGlyValProLysValIleGlySerAsp 957
Qy      2758 AATGGTCCAGCTTTTGTGCTTGGTCCAGTAAGTACGAGGACTGGCCAGATATTGGGATTCAT 2817
Db      958 AsnGlyProAlaPheValAlaGlnValSerGlnGlyLeuAlaThrGlnLeuGlyIleAsp 977
Qy      2818 TGGAACTGCAATTTGTCATACAGACCCCAAGCTCAGACAGGTAGAGAGATCAATAGA 2877
Db      978 TrpLysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArg 997
Qy      2878 ACATTAAGAGACCTTACTAATTGACCGCGGAGACTGCGGTTAATGATGGATAGCT 2937
Db      998 ThrLeuLysGluThrLeuThrLysLeuAlaMetGluThrGlyLysAspTrpValAla 1017
Qy      2938 CTCTGCGCTTTGTGCTTTTCCAGCTTTTCTCTAGGCTCAAGGCACTTGAGTGGGTG 3051
Db      1018 LeuLeuProLeuAlaLeuPheArgAlaArgAsnThrProGlyArgPheGlyLeuThrPro 1037
Qy      2998 TATGAATTAATCTACGGGGGACCCCGCCCATTTGGTAGAAATTTGTTCT-----GTACAT 3051
Db      1038 PheGluValLeuTyrGlyGlyProProProLeuIleLysAspGlyGlyThrLeuValPro 1057
Qy      3052 AGTGCTGACGTGCTTTTCCAGCTTTTCTCTAGGCTCAAGGCACTTGAGTGGGTG 3111
Db      1058 AspSerGlySerValLeuProSerSerLeuLeuIleHisLeuLysAlaLeuLysValIle 1077
Qy      3112 AGACACGAGCGGTGGAGGCACTCCGGAGCGCTTACTCAGAGGAGGAGACTTCAGATC 3171
Db      1078 ArgThrGlnIleTrpAspGlnLeuLysThrAlaTyrThr---ProGlyThrThrAlaVal 1096
Qy      3172 CCACATCGTTTCAAGTGGGAGATTCACTCTACGTTAGACGCCACCGTGCAGGAACCTC 3231
Db      1097 ProHisGluPheGlnValGlyAspGlnValLeuValArgArgHisArgThrGlySerLeu 1116
Qy      3232 GAGACTCGGTGGAGGCGCTTATCTCGTACTTTTGTACCAACACCAACCGCTGGAAGTC 3291
Db      1117 GluProArgTrpLysGlyProTyrLeuValLeuLeuThrThrProThrAlaValLysVal 1136
Qy      3292 GAAGGATCTCCACCTGGATCCATGATCCAGTTAAACGGGGCGGCACCTCCCGAT--- 3348
Db      1137 AspGlyIleAlaSerTrpIleHisAlaSerHisValLysArgAlaProSerGlnAspGlu 1156
```

```
Qy      3349 -----TCGGGTGGAAAGCCGAAAGACTGAAATCCCTTAAGCTTCGCCTC 3396
Db      1157 GluThrHisGluAspAsnTrpAlaValGluAlaThrAspAsnProLeuLysLeuArgLeu 1176
Qy      3397 CATCGCTGGTTCTCT 3411
Db      1177 LeuArgArgSerPro 1181
RESULT 3
US-09-309-572-13
; Sequence 13, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 1737
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURES:
; OTHER INFORMATION: gag-pol protein
US-09-309-572-13
Alignment Scores:
Pred. No.: 0 Length: 1737
Score: 3797.50 Matches: 719
Percent Similarity: 76.3% Conservative: 165
Best Local Similarity: 62.0% Mismatches: 242
Query Match: 61.4% Indels: 33
DB: 2 Gaps: 12
US-10-723-552-3_COPY_2307_5741 (1-3435) x US-09-309-572-13 (1-1737)
```

```
Qy      4 GTTGCCACAGGGCAACACAGTATCCATGACTACCCGAAAGAACAGTTGATGGAGTG 63
Db      589 GlyAlaThrGlyGlyLysArgTyrArgTyrThrAspArgLysValHisLeuAlaThr 608
Qy      64 GGACGGGTAAACCCACCTGTTTCTGGTCATACCTGAGTGGCCAGCACCCCTCTTAGGTAGA 123
Db      609 GlyLysValThrHisSerPheLeuHisValProAspCysProTyrProLeuLeuGlyArg 628
Qy      124 GACTTATTGACCAAGATGGGAGCACAAATTTCTTTTGA---CAAGGGAACCAAGAGTG 180
Db      629 AspLeuLeuThrLysLysLysAlaGlnIleHisPheGluGlySerGlyAlaGlnValMet 648
Qy      181 TCTGCAATAACAAACCTATCACTGTGTGACCTCCCAATTAGATGACGAATATCGACTA 240
Db      649 GlyProMetGlyGlnProLeuGlnValLeuThrLeuAsnIleGluAspGluHisArgLeu 668
Qy      241 TACTCTCCCTAGTAAGCTGATCAAAATATA---CAATTCTGTTGGAACAGTTTCCC 297
Db      669 HisGluThrSerLysGluProAspValSerLeuGlySerThrTrpLeuSerAspPhePro 688
Qy      298 CAAGCTGGGCAACAAACCCGAGGATGGTTTGGCAAGCAAGTCTCCCCACCAAGTTATT 357
Db      689 GlnAlaTrpAlaGluThrGlyGlyMetGlyLeuAlaValArgGlnAlaProLeuIleIle 708
Qy      358 CAATCAAGCCAGTCCACACAGTGTCTCAGTCAGACAGTACCCCTTGTAGTAAGAAGCT 417
Db      709 ProLeuLysAlaThrSerThrProValSerIleLysGlnTyrProMetSerGlnGluAla 728
Qy      418 CAAGAAGGATTCGGCGGATGTCAAAGATTAAATCAACAGGCGCATCTTAGTTCCTGTC 477
Db      729 ArgLeuGlyIleLysProHisIleGlnArgLeuAspGlnGlyIleLeuValProCys 748
```

QY 478 CAATCTCCCTGGAATACCTCCCTGCTACCGGTTAGAAAGCCCTGGAGCTAAATGACTATCGA 537  
DB 479 GlnSerProThrPasnThrProLeuLeuProValLysLysProGlyThrPasnAspLysArg 768  
QY 538 CCAGTACAGGACTTTGAGAGAGGTCAATAAACCAGGTGCGAGTATACACCCCAACAGTCCCG 597  
DB 539 ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro 788  
QY 598 AACCTTATACCTCTTGTGTGCTCTCCACCCCAACCGAGCTGGTATACAGTATTGGAC 657  
DB 599 AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp 808  
QY 658 TTAAGGATGCTCTCTCTGCTGAGATTACACCCCTAGCCCAACCACTTTTTCCTTC 717  
DB 659 LeuLysAspAlaPhePheCysLeuArgLysProThrSerGlnProLeuPheAlaPhe 828  
QY 718 GAATGGAGAGATCCAGGTACCGGAGAACCCCGGAGCTCACCTGACCCGAGTCCCCCAA 777  
DB 719 GluTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln 848  
QY 778 GGGTTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACAGAGAGCTGGCCAACTTC 837  
DB 779 GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe 868  
QY 838 AGGATCCCAACACCTCCAGTGCACCTCTCCAGTACGTGGATGACCTGCTTCTGCGGGA 897  
DB 839 ArgIleGlnHisProAspLeuLeuLeuLeuGlnTrpValAspAspLeuLeuAlaAla 888  
QY 898 GCCCAACCAAGGACTGCTTGAAGGACACGAGGCACTACTGCTGGAAATGCTGCACCTA 957  
DB 899 ThrSerGluLeuAspCysGlnGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAsnLeu 908  
QY 958 GGCTACAGAGCTCTGTGAAGGCCAGATTTGACGAGAGAGAGGTAACTACTTGGGG 1017  
DB 959 GlyTyrArgAlaSerAlaLysLysAlaGlnIleCysGlnLysGlnValLysTyrLeuGly 928  
QY 1018 TACAGTTTGGGAGGCGGACGAGTGGCTGACGAGGACCGAGAACTGTAGTCCAG 1077  
DB 929 TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGlnThrValMetGly 948  
QY 1078 ATACGGGCCCCAACCAACCAACCAACCAATGAGAGAGTGTGGGACAGCTGGATTTGC 1137  
DB 949 GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys 968  
QY 1138 AGACTGTGGATCCCGGGTTTGGACCTTACGAGCCCACTTACCCCTAACCAAGAA 1197  
DB 969 ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr 988  
QY 1198 AAAGGGGAATTCCTCGGCTCCTGAGCACAGAGGCAATTTGATGCTATCAAAAAGGCC 1257  
DB 989 GlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTyrGlnGluIleLysGlnAla 1008  
QY 1258 CTGCTGAGCGCACTGCTCGGCCCTCCCTGACGTAACTAAACCCCTTTACCCCTTTATGTG 1317  
DB 1009 LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal 1028  
QY 1318 GATGAGCTAAGGAGTAGCCCGGGAGTGTAAACCCAAACCTTAGGACCATGAGAGA 1377  
DB 1029 AspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg 1048  
QY 1378 CCTGTCGCTACCTGTCAAGAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCCCTG 1437  
DB 1049 ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTrpProCysLeu 1068  
QY 1438 AAGGCTACGAGCTGTGGCCATCTGCTCAAGAGCGCTGACAAATTCACCTTTGGGACAG 1497  
DB 1069 ArgMetValAlaAlaIleAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGln 1088  
QY 1498 AATTAACGTAAATGACCCCATGATGAGAGAACATGTTGGGACGCCCCACGACCGA 1557  
DB 1089 ProLeuValIleLeuAlaProHisAlaValGluAlaLeuValLysGlnProProAspArg 1108  
QY 1558 TGGATGCCAACCGCCGACATGACCCCACTATCAAGCCTGCTCTC--ACAGAGAGGGTC 1614

DB 1109 TrpLeuSerAsnAlaArgMetThrHisTyrGlnAlaLeuLeuAspThrAspArgVal 1128  
QY 1615 AGTTTCGCTCCACAGCGGCTCTCAACCTGCGCACTCTTCTGCTGAGAGACTGATGAA 1674  
DB 1129 GlnPheGlyProValValAlaLeuAsnProAlaThrLeuLeuPro--LeuProGluGlu 1147  
QY 1675 CCAGTGAATCATGATGCCCATCACTATTGATTCAGGAGACTGGGGTCCGCAAGACCTT 1734  
DB 1148 GlyLeuGlnHisAsnCysLeuAspIleLeuAlaGluAlaHisGlyThrArgProAspLeu 1167  
QY 1735 ACAGACATACCGCTGACTGGAGAAAGTCTAACCTGGTTCTACCTGACGGAAGCAGTATGTG 1794  
DB 1168 ThrAspGlnProLeuProAspAlaAspHisThrTrpTyrThrAspGlySerSerLeuLeu 1187  
QY 1795 GTGAAGGTAAAGAGATGGCTGGGGCGGTGTGTGACGGGACCGGCACCATCTGGGCC 1854  
DB 1188 GlnGluGlyGlnArgLysAlaGlyAlaAlaValThrThrGluThrGluValIleTrpAla 1207  
QY 1855 AGCAGCTGCGGGAAGGAACCTTCAGCACAAAGGCTGAGCTCATGGCCCTCACCAAGCT 1914  
DB 1208 LysAlaLeuProAlaGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAla 1227  
QY 1915 TTGGCGCTGGCCGGAAGGAATCCATAAATTTATACGACAGCAGGTATGCTTTTTCG 1974  
DB 1228 LeuLysMetAlaGluGlyLysLysLeuAsnValTyrThrAspSerArgTyrAlaPheAla 1247  
QY 1975 ACTCACAGCTACATGGGGCCATCTATAACAAGGGGTTGCTTACTCTAGCAGGAGG 2034  
DB 1248 ThrAlaHisIleHisGlyGluIleTyrArgArgGlyLeuLeuThrSerGluGlyLys 1267  
QY 2035 GAAATAAGACAAGAGGAATTTCTAAGCTATTAGAACGCTACATCTCATATACCAAAAAGG 2094  
DB 1268 GluIleLysAsnLysAspGluIleLeuAlaLeuLysAlaLeuPheLeuProLysArg 1287  
QY 2095 CTAGCTATTATACACTGCTCTGGACATCAGAAAAGCTAAAGATCTCATATACAGAGGAAC 2154  
DB 1288 LeuSerIleIleHisCysProGlyHisGlnLysGlyHisSerAlaGluAlaArgGlyAsn 1307  
QY 2155 CAGATGCTGACCGGTTGCCAAGCAGCAGCCCGGGTGTAACTTCTGCTATATA 2214  
DB 1308 ArgMetAlaAspGlnAlaAlaArgLysAlaAla-----IleThr 1320  
QY 2215 GAAATGCCCAAGGC-----CCAGAACCACAGA 2241  
DB 1321 GluThrProAspThrSerThrLeuLeuIleGluAsnSerSerProTyrThrSerGluHis 1340  
QY 2242 CGACAGTACACCTTAGAAGACTGGCAAGAGATAAAGATAGACCAG---TTCTCTGAG 2298  
DB 1341 PheHisTyrThrValThrAspIleLysAspLeuThrLysLeuGlyAlaIleTyrAspLys 1360  
QY 2299 ACTCCGAGAGGAGCTGCTATACCTCAGATGGGAAGAAATCTGCGCCCAACAAAGAGG 2358  
DB 1361 ThrLysLysTyrTrpValTyr-----GlnGlyLysProValMetProAspGlnPheThr 1378  
QY 2359 TTAGAATATGCCACAGATACATCGTCTAACCCACCTAGGAACCTAAACACCTGCAGCAG 2418  
DB 1379 PheGluLeuLeuAspPhePheLeuHisGlnLeuThrHisLeuSerPheSerLysMetLysAla 1398  
QY 2419 TTGGTTCAGAAC-----TCCCTTATCATCTTCTGAGGCTTACCAGGAGTGGCTGACCTCG 2472  
DB 1399 LeuLeuGluArgSerHisSerProTyrTyrMetLeuAsnArgAspArgThrLeuLysAsn 1418  
QY 2473 GTGTCAACAACATGTGTGCCCTGCCAGCTGTTTAATGCTAATCTTCCAGAAATGCTCCA 2532  
DB 1419 IleThrGluThrCysLysAlaCysAlaGlnValAsnAlaSerLysSerLysValLysGln 1438  
QY 2533 GGGAGAGACTAAGGGGAGCCACCCAGGCGCTCACTGGGAGTGGACTTCACTGAGTA 2592  
DB 1439 GlyThrArgValArgGlyHisArgProGlyThrHisTrpGluIleAspPheThrGluIle 1458  
QY 2593 AAGCGGCTAAATACGGGAACAAATACCTATGTTGTTTTGTAGACACCTTTTTCAGGATGG 2652

Db 1459 LysProGlyLeuTyrGlyTyrLysTyrLeuLeuValPheIleAsePThrPheSerGlyTrp 1478  
Qy 2653 GTAGAGGCTTACTCTAAGAAAGAGACTTCAACCGGTGGCTAAATAAATACTGAA 2712  
Db 1479 IleGluAlaPheProThrLysLysGluThrAlaLysValValThrLysLysLeuGlu 1498  
Qy 2713 GAAATTTTCCAGATTTGGAATACCTAAGTAATAGGTCAGACAATAGTCCAGCTTTT 2772  
Db 1499 GluIlePheProArgPheGlyMetProGlnValLeuGlyThrAspAsnGlyProAlaPhe 1518  
Qy 2773 GTTGGCCAGTAAGTCAGGACTGCGCAAGATATTGGGATTTGATTTGAAATCTGATGT 2832  
Db 1519 ValSerLysValSerGlnThrValAlaAspLeuGlyIleAspTrpLysLeuHisCys 1538  
Qy 2833 GCATACAGACCCCAAGCTCAGGACAGGTAGAGAGATCAATAGAACCATTAAGAGACC 2892  
Db 1539 AlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThr 1558  
Qy 2893 CTTACTAATTTGACCGCGAGACTGCGCTTAATGATTTGGATAGCTCTCTCGCTTTGTG 2952  
Db 1559 LeuThrLysLeuThrLeuAlaThrGlySerArgAspTrpValLeuLeuProLeuAla 1578  
Qy 2953 CTTTGTAGGTTAGGAACACCCCTGGACAGTTTGGCTGACCCCTATGAATTAATCTTAC 3012  
Db 1579 LeuTyrArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGluIleLeuTyr 1598  
Qy 3013 GGGGACCCCTCCATTCGTAGAAATTCCTTCTGTACATAGTGTGCTGCTGCTGCTTCC 3072  
Db 1599 GlyAlaProProProLeuValAsnPheProAspProAspMetThrArgValThrAsnSer 1618  
Qy 3073 CAGCTTTGCTCTAGGCTCAAGGCACTTGAAGTGGTGAGACCAACGAGCGTGGAGCAA 3132  
Db 1619 ProSerLeuGlnAlaHisLeuGlnAlaLeuTyrLeuValGlnHisGluValTyrArgPro 1638  
Qy 3133 CTCGGGAGGCTACTCAGGAGGAGGAGCTTGCAG---ATCCACATCGTTTCCAAGTG 3189  
Db 1639 LeuAlaAlaTyrGlnGlnLeuAspArgProValValProHisProTyrArgVal 1658  
Qy 3190 GGAGATTCACTAGTTAGACGCCACCGTCAGGAAACCTCGAGACTCGGTGGAAGGGC 3249  
Db 1659 GlyAspThrValTyrValArgArgHisGlnThrLysAsnLeuGluProArgTrpLysGly 1678  
Qy 3250 CTTATCTCGTACTTTGACACACCAACCGCTGTGAAGTCGAAGGAATCTCCACCTGG 3309  
Db 1679 ProTyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleAlaAlaTyr 1698  
Qy 3310 ATCCATGCATCCACGTTAAACCGCGCCACT-----CCGATTTCG-----GGG 3354  
Db 1699 IleHisAlaAlaHisValLysAlaAlaAspProGlyGlyGlyProSerSerArgLeuThr 1718  
Qy 3355 TGGAAAGCCGAAAGACTGAAATCCCTTAAGCTTCGCTCCATCGCTGCTGTTCCCT 3411  
Db 1719 TrpArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGluAlaPro 1737

## RESULT 4

US-09-718-096-13  
; Sequence 13, Application US/09718096  
; Patent No. 6589763  
; GENERAL INFORMATION:  
; APPLICANT: Von Leer, Meike-Dorothee  
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV  
; FILE REFERENCE: 35-195  
; CURRENT APPLICATION NUMBER: US/09718,096  
; CURRENT FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: DE 19856463.5  
; PRIOR FILING DATE: 1998-11-26  
; PRIOR APPLICATION NUMBER: EP 99250415.9  
; PRIOR FILING DATE: 1999-11-25  
; PRIOR APPLICATION NUMBER: US 09/309,572  
; PRIOR FILING DATE: 1999-05-11  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13

; LENGTH: 1737

; TYPE: PRT

; ORGANISM: Moloney murine leukemia virus

; FEATURE:

; OTHER INFORMATION: gag-pol protein

US-09-718-096-13

## Alignment Scores:

Pred. No.:	0	Length:	1737
Score:	3797.50	Matches:	719
Percent Similarity:	76.3%	Conservative:	165
Best Local Similarity:	62.0%	Mismatches:	242
Query Match:	61.4%	Indels:	33
DB:	2	Gaps:	12

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-09-718-096-13 (1-1737)

Qy 4 GGTGCCACAGGCAACACAGTATCCATGGACTACCCGAAAGAACAGTTGACTTGGAGTG 63  
Db 589 GlyAlaThrGlyGlyLysArgTyrArgTrpThrThrAspArgLysValHisLeuAlaThr 608  
Qy 64 GGACGGGTAAACCACTCGTTTCTCGTCATACCTGAGTGCCCGACACCCCTCTTAGGTAGA 123  
Db 609 GlyLysValThrHisSerPheLeuHisValProAspCysProTyrProLeuLeuGlyArg 628  
Qy 124 GACTTATTGACCAAGATGGGAGCACAAATTTCTTTTGA---CAAGGGAAACACAGAGTG 180  
Db 629 AspLeuLeuThrLysLeuLysAlaGlnIleHisPheGluGlySerGlyAlaGlnValMet 648  
Qy 181 TCTGCAAAATCAAACTCATCTGTTGACCTCCAAATTTAGATGACGAATATCGACTA 240  
Db 649 GlyProMetGlyGlnProLeuGlnValLeuThrLeuAsnIleGluAspGluHisArgLeu 668  
Qy 241 TACTCTCCCTAGTAAAGCCTGATCAAAATATA---CAATTCTGTTGGACAGTTTCCC 297  
Db 669 HisGluThrSerLysGluProAspValSerLeuGlySerThrTrpLeuSerAspPhePro 688  
Qy 298 CAACCTCGGCGAGAAACCGCAGGATGGTTTGGCAAGCAAGTTCCTCCCAACAGTTATT 357  
Db 689 GlnAlaTrpAlaGluThrGlyGlyMetGlyLeuAlaValArgGlnAlaProLeuIleIle 708  
Qy 358 CAACTGAAGCCAGTGCACACCACTGTCAGTCAAGTACAGTACCTCTTGAGTAAGAAGCT 417  
Db 709 ProLeuLysAlaThrSerThrProValSerIleLysGlnTyrProMetSerGlnGluAla 728  
Qy 418 CAAGAAAGAAATTCGGCCGATGTCCTCAAGATTAATCCAAAGGCGATCTTGTCTCTC 477  
Db 729 ArgLeuGlyIleLysProHisIleGlnArgLeuLeuAspGlnGlyIleLeuValProCys 748  
Qy 478 CAATCTCCCTGGAAATCTCCCTGCTACCGTTAGAAAGCTGGGACTTAATGACTATCGA 537  
Db 749 GlnSerProTrpAsnThrProLeuLeuProValLysLysProGlyThrAsnAspTyrArg 768  
Qy 538 CCAGTACAGACTTGAGAGAGGTCAATAAACGGGTGCGAGGATATACACCAAGTCCCG 597  
Db 769 ProValGlnAspLeuArgGluValAsnLysArgValGluAspIleHisProThrValPro 788  
Qy 598 AACCTTATAACCTCTTGTGTCTCTCCACCCCAACCGAGCTGGTATACAGTATTGGAC 657  
Db 789 AsnProTyrAsnLeuLeuSerGlyLeuProProSerHisGlnTrpTyrThrValLeuAsp 808  
Qy 658 TTAAGGATGCTTCTCTGCTGAGATTACACCCCACTAGCCCAACCACTTTTTCCTTC 717  
Db 809 LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPhe 828  
Qy 718 GAATGGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACTGGACCCGACCTGCCCAA 777  
Db 829 GluTrpArgAspProGluMetGlyIleSerGlyGlnLeuThrTrpThrArgLeuProGln 848  
Qy 778 GGGTTCAAGAACTCCCGGACCATCTTTGACGAAGCCCTACACAGACCTGGCCCACTTC 837  
Db 849 GlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe 868



QY 838 AGGATCCAAACACCTCAGGTGACCTCTCCAGTACGTGGATGACCTGCTCTCGCGGA 897  
Db ArgilleGlnHisProAspLeuLeuLeuGlnThrValAspAspLeuLeuLeuAla 888  
QY 898 GCCACCAACAGGACTGCTTAGAAGGCACGAGGACCTACTGCTGGAATTCCTGACCTA 957  
Db ThrSerGluLeuAspCysGlnGlyThrArgAlaLeuLeuGlnThrLeuGlyAsnLeu 908  
QY 958 GGCTACAGAGCTCTGCTAAGAGCCCGAGATTTGCGAGGAGAGAGTAACTACTTGGGG 1017  
Db GlyTyrArgAlaSerAlaLysAlaGlnIleCysGlnLysGlnValLysThrLeuGly 928  
QY 1018 TACAGTTTGGCGGACGGCAGCGATGGCTGACGAGGACCGAGAAACCTGTAGTCCAG 1077  
Db TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGlnThrValIleGly 948  
QY 1078 ATACCGGCCCCAACCCACGCAACAAATGAGAGAGATTTTGGGACAGCTGGATTTGTC 1137  
Db GlnProThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCys 968  
QY 1138 AGACTGTGGATCCCGGGTTTGGACCTTAGCAGCCCGACCTACTACCGCTAACCAAGAA 1197  
Db ArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr 988  
QY 1198 AAAGGGGAATTCCTCGGGCTCTGAGCACCAGAGGCAATTTGATGCTATCAAAAAGGCC 1257  
Db GlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTrpGlnGluLeuGlnAla 1008  
QY 1258 CTGCTGAGCGCACCTGCTCTGGCCCTCCTGACGTAACCTTATACCTTTATGTTG 1317  
Db LeuLeuThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheVal 1028  
QY 1318 GATGAGCTAAGGAGTACCGGGGAGTTTAAACCCAAACCTTAGGACCATGAGAGA 1377  
Db AspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg 1048  
QY 1378 CCTGCGCTACCTGTCAAAGAGCTCGATCCTGTAGCCAGTGGTGGCCCATATGCTG 1437  
Db ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTrpProCysLeu 1068  
QY 1438 AAGGCTATCGCAGCTGTGGCCATCTGCTCAAGGACGCTGACAAATGACTTTGGGACAG 1497  
Db ArgMetValAlaAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGln 1088  
QY 1498 AATATACTGTAATAGCCCCCATGCTGATGAGAACATCGTTGGCAGCCCCCAGACCGA 1557  
Db ProLeuValIleLeuAlaProHisAlaValGluAlaLeuValLysGlnProProAspArg 1108  
QY 1558 TGGATGACCAACCGCCGATGACCCACCTATCAAGCCCTGCTTCTC--ACAGAGAGGCTC 1614  
Db TrpLeuSerAsnAlaArgMetThrHisTyrGlnAlaLeuLeuLeuAspThrAspArgVal 1128  
QY 1615 ACGTTTCGCTCCACCGCGCTCTCAACCCCTGCACTCTTCTGCTGGAAGACACTGATGA 1674  
Db GlnPheGlyProValValAlaLeuAsnProAlaThrLeuLeuPro--LeuProGluGln 1147  
QY 1675 CCAGTGACTCATGATGTCATCAACTATGATGAGGAGACTGGGGTCCGCAAGACCTT 1734  
Db GlyLeuGlnHisAsnCysLeuAspIleLeuAlaGlnAlaHisGlyThrArgProAspLeu 1167  
QY 1735 ACAGACATACCGCTGACTGAGGAGTCTAACCTGTTTCACTGACGGAAGCAGCTATGTCG 1794  
Db ThrAspGlnProLeuProAspAlaAspHisThrTrpTyrThrAspGlySerSerLeuLeu 1187  
QY 1795 GTGGAAGGTAAGAGGATGGCTGGGGCGCGGTGGTGGAGCGGACCCGACCATCTGGGCC 1854  
Db GlnGluGlyGlnArgLysAlaGlyAlaAlaValThrThrGluThrGluValIleTrpAla 1207  
QY 1855 AGCAGCCTCGCGGAAGAACTTCAGCACAAAGGCTGAGCTCATGGCCCTCAGCGAAGCT 1914  
Db LysAlaLeuProAlaGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAla 1227  
QY 1915 TTGCGGCTGGCGAAGGGAATCCATAACATTTTATACGGACAGCAGGTATGCTTTCGG 1974

Db LeuLysMetAlaGluGlyLysLeuAsnValTyrThrAspSerArgTyrAlaPheAla 1247  
QY 1975 ACTGCACACGTACATGGGCCATCTATAACAAGGGGGTTCCTACTCTCAGCAGGAGG 2034  
Db ThrAlaHisIleHisGlyGluIleTyrArgArgGlyLeuLeuThrSerGluGlyLys 1267  
QY 2035 GAAATAAGAACAAAGAGGAAATTTCTAAGCTATTAGAGCCGTACATTTTACCAAAAAGG 2094  
Db GluIleLysAsnLysAspGluIleLeuAlaLeuLeuLysAlaLeuPheProLysArg 1287  
QY 2095 CTAGCTATTATACACTCTCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAAC 2154  
Db LeuSerIleIleHisCysProGlyHisGlnLysGlyHisSerAlaGluAlaArgGlyAsn 1307  
QY 2155 CAGATGCTGACCGGGTTCGCAAGCAGGACCCAGGCTGTAACTTCTGCTCTATATA 2214  
Db ArgMetAlaAspGlnAlaAlaArgLysAlaAla-----IleThr 1320  
QY 2215 GAATGCCCAAGGCC-----CCAGAACCACG 2241  
Db GluThrProAspThrSerThrLeuLeuIleGluAsnSerSerProTyrThrSerGluHis 1340  
QY 2242 CGACAGTACACCCCTAGAAGACTGGCAAGAGATAAAAAGATAGACCCAG---TTCTCTGAG 2298  
Db PheHisTyrThrValThrAspIleLysAspLeuThrLysLeuGlyAlaIleTyrAspLys 1360  
QY 2299 ACTCCGGAAGGACCTCTCTATACCTCAGATGGGAGGAAATCTCGCCCCCAAGAGAGG 2358  
Db ThrLysLysTyrTrpValTyr-----GlnGlyLysProValMetProAspGlnPheThr 1378  
QY 2359 TTAGAATATGTCACACAGATACATCGTCAACCCACCTAGGAACCTAAACACCTCGACGAG 2418  
Db PheGluLeuLeuAspPheLeuHisGlnLeuThrHisLeuSerPheSerLysMetLysAla 1398  
QY 2419 TTGCTCAGAACCA-----TCCCTTATCATGTTCTGAGGCTACACGAGTGGCTGACGTCG 2472  
Db LeuLeuGluArgSerHisSerProTyrTyrMetLeuAsnArgAspArgTyrLeuLysAsn 1418  
QY 2473 GTGCTCAAACTTGTGTGCTGCGCCTGCTGCTGTTTAACTGCTAATCTCTCAGAAATGCCTCCA 2532  
Db IleThrGluThrCysLysAlaCysAlaGlnValAsnAlaSerLysSerAlaValLysGln 1438  
QY 2533 GGAAGAGACTTAAGGGAAGCCACCCAGGCGCTCACTGGGAAGTGACTCACTGAGGTA 2592  
Db GlyThrArgValArgGlyHisArgProGlyThrHisTrpGluIleAspPheThrGluIle 1458  
QY 2593 AAGCGGCTAAATACGGAACCAATACCTATTGTTGTTTGTAGACACCTTTTCAGGATGG 2652  
Db LysProGlyLeuTyrGlyTyrLysTyrLeuLeuValPheIleAspThrPheSerGlyTrp 1478  
QY 2653 GTAGAGGCTTATCCTACTAAGAAAGAGACTTCAACCGTGGTGGCTTAAATAATACTGAA 2712  
Db IleGluAlaPheProThrLysLysGluThrAlaLysValValThrLysLysLeuGlu 1498  
QY 2713 GAATTTTCCAAAGATTGGAAATACCTAAGCTAATAGGTCAGCAATGTGCTCCAGCTTTT 2772  
Db GluIlePheProArgPheGlyMetProGlnValLeuGlyThrAspAsnGlyProAlaPhe 1518  
QY 2773 GTTCCCGAGGTAAAGTTCAGGGAAGTGGCCCAAGATATTGGGGATTGATTGGAATCGATTGT 2832  
Db ValSerLysValSerGlnThrValAlaAspLeuLeuGlyIleAspTrpLysLeuHisCys 1538  
QY 2833 GCATACAGACCCCAAGCTCAGGACAGGTAGAGGATGAATAGAACCTTAAGAGAGCC 2892  
Db AlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThr 1558  
QY 2893 CTTACTAAATTGACCGCGGAGACTGGCGTTTAACTGATGATGCTCTCTGCTCCCTTTCGTG 2952  
Db LeuThrLysLeuThrLeuAlaThrGlySerArgAspTrpValLeuLeuLeuProLeuAla 1578  
QY 2953 CTTTTTAGGTTAGGAACACCCCTCGACAGTTTGGGCTGACCCCTATGAATTTACTCTTAC 3012

Db 1579 LeuTyrArgAlaAraAenThrProGlyProHisGlyLeuThrProTyrGluLeuLeuTyr 1598  
Qy 3013 GGGGACCCCCCATTTGGTAGAAATTCCTCTGTACATAGTGTGACGTGCTCTTCC 3072  
Db 1599 GlyAlaProProLeuValAenPheProAspMetThrArgValThrAsnSer 1618  
Qy 3073 CAGCCTTTGCTCTAGGCTCAAGGACCTTGGTGGGTGAGACAAACAGCGGTGAGGCAA 3132  
Db 1619 ProSerLeuGlnAlaHisLeuGlnAlaLeuTyrLeuValGlnHisGluValTTPArgPro 1638  
Qy 3133 CTCGGGAGGCGCTACTCAGGAGGAGGAGCTTGCAG---ATCCACACATGTTTCCAAAGTG 3189  
Db 1639 LeuAlaAlaAlaTyrGlnGlnLeuAaspArgProValProHisProTyrArgVal 1658  
Qy 3190 GGAGATTCACTTACGTTAGACGCCACCGTGCAGAAACCTCGAGACTCGGTGGAAGGC 3249  
Db 1659 GlyAaspThrValTTPValArgArgHisGlnThrLysAsnLeuGluProAsgTTPLeuGly 1678  
Qy 3250 CCTTATCTCGTACTTTGACACACCAACCGCTGTGAAGTCGAAAGTCTCCACCTGG 3309  
Db 1679 ProTyrThrValLeuLeuThrProThrAlaLeuLysValAaspGlyIleAlaAlaTTP 1698  
Qy 3310 ATCCATCATCCACGTTTAAACCGCGCCACCT-----CCGATTCCG-----GGG 3354  
Db 1699 IleHisAlaAlaHisValLysAlaAlaAaspProGlyGlyProSerSerArgLeuThr 1718  
Qy 3355 TGGAAACCCGAAAGACTGAAATCCCTTTAAGCTTCGCTCCATCCGCTGTTCCCT 3411  
Db 1719 TTPArgValGlnArgSerGlnAenProLeuLysIleArgLeuThrArgGluAlaPro 1737

RESULT 5

US-08-929-967-8  
; Sequence 8, Application US/08929967  
; Patent No. 5891637  
; GENERAL INFORMATION:  
; APPLICANT: Rupert, Siegfried J.W.  
; TITLE OF INVENTION: Construction of Full-Length cDNA Libraries  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/929,967  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: P1035R1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-3216  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1079 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
US-08-929-967-8  
Alignment Scores:  
Pred. No.: 0  
Score: 3600.00  
Percent Similarity: 76.7%  
Best Local Similarity: 62.4%  
Length: 1079  
Matches: 679  
Conservative: 156  
Mismatches: 222

Query Match: 58.2% Indels: 32  
DB: 1 Gaps: 11  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-08-929-967-8 (1-1079)  
Qy 211 ACCCTCAATTAGATGACGAATATCGACTATATCTCTCCCTAGTAAAGCTGATCAAAAT 270  
Db 1 ThrLeuAsnIleGluAaspGluHisArgLeuHisGluThrSerLysGluProAaspValSer 20  
Qy 271 ATA--CAATTCGTGGTGAACAGTTTCCCAAGCCTGGGCAGAAAACCCGAGGATGGGT 327  
Db 21 LeuGlySerThrTTPLeuSerAaspPheProGlnAlaTTPAlaGluThrGlyGlyMetGly 40  
Qy 328 TTGCGAAGCAAGTTCCCCCAAGTTTATTCATCTGAAGCGCCAGTGCACACCACTGTGCA 387  
Db 41 LeuAlaValArgGlnAlaProLeuIleLeuProLeuLysAlaThrSerThrProValSer 60  
Qy 388 GTCAGACAGTACCCCTTGAGTAAAGAGCTCAAGAGAAATTCGGCGCATGTCCAAGA 447  
Db 61 IleLysGlnTyrProMetSerGlnGluAlaArgLeuGlyIleLysProHisIleGlnArg 80  
Qy 448 TTAATCCAAACAGGSCATCCTAGTTCCTGTCCCAATCTCCCTGGAATACTCCCTGCTACCG 507  
Db 81 LeuLeuAaspGlnGlyIleLeuValProCysGlnSerProTTPAsnThrProLeuLeuPro 100  
Qy 508 GTTAGAAAGCCTGGGACTAATGACTATCGACACAGTACAGGACTTGAGAGAGGTCAATAAA 567  
Db 101 ValLysLysProGlyThrAsnAaspTyrArgProValGlnAaspLeuArgGluValAsnLys 120  
Qy 568 CGGTGTCAGATATACACCCAAACAGTCCCGAACCTTTATACCTCTTGTGTCTCTCCA 627  
Db 121 ArgValGluAaspIleHisProThrValProAaspProTyrAsnLeuLeuSerGlyLeuPro 140  
Qy 628 CCCCACCGAGCTGGTATACAGTATTCGACTTAAAGGATCCCTCTCTCTCCCTGAGATTA 687  
Db 141 ProSerHisGlnTTPThrValLeuAaspLeuLysAaspAlaPhePheCysLeuArgLeu 160  
Qy 688 CACCCACTAGCCAAACACCTTTTTCCTTCGAATGGAGAGATCCAGGTACGGGAGGAAC 747  
Db 161 HisProThrSerGlnProLeuPheAlaPheGluTTPArgAaspProGluMetGlyIleSer 180  
Qy 748 GGCAGCTACCTGGACCCGACTGCGCCCAAGGTTCAAGAACTCCCGACCATCTTTTGCAC 807  
Db 181 GlyGlnLeuThrTTPThrArgLeuProGlnGlyPheLysAasnSerProThrLeuPheAasp 200  
Qy 808 GAAGCCCTACACAGACACCTGGCCCACTTCAGATCCCAACACCTCAGGTGACCTCTC 867  
Db 201 GluAlaLeuHisArgAaspLeuAlaAaspPheArgIleGlnHisProAaspLeuIleLeu 220  
Qy 868 CAGTACGTGGATGACCTGCTTCTGGCGGGAGCCACCAACACAGGACTGCTTAGAAGGCACG 927  
Db 221 GlnTyrValAaspAaspLeuLeuAlaAlaThrSerGlnLeuAaspCysGlnGlnGlyThr 240  
Qy 928 AAGGCACTACTGCTGAAATGCTCTGACCTAGGCTCAGAGCCTCTGCTAGAAGGCCACG 987  
Db 241 ArgAlaLeuLeuGlnThrLeuGlyAasnLeuGlyTyrArgAlaSerAlaLysAlaGln 260  
Qy 988 ATTTCAGGAGAGAGGTAACTACTTGGGTACAGTTTGGCGGACCGGCGGACCGATGGCTG 1047  
Db 261 IleCysGlnLysGlnValLysTyrLeuGlyTyrLeuLeuLysGluGlyGlnArgTTPLeu 280  
Qy 1048 ACGGAGGCACGGAAGAAAACCTGTAGTCCAGATACCGGCCCAACACAGCAACAATAATG 1107  
Db 281 ThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGlnLeu 300  
Qy 1108 AGAGATTTTTTGGGACAGCTGGATTTTGCAGACTGTGAGTCCCGGGTTTGGCACTTA 1167  
Db 301 ArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTTPLeuPheAlaGluMet 320  
Qy 1168 GCAGCCCACTCTACCGGCTAACCAAGAAAAGGGGAATTCCTCTGGGCTCTCTGACAC 1227  
Db 321 AlaAlaProLeuTyrProLeuThrLysThrGlyThrLeuPheAasnTTPGlyProAaspGln 340









Db	425	ArgGluThrGlnLysAlaAsnThrTyrLeuValArgTrpThrProThr---	GluValAla	443
Qy	1237	TTTTCATGCTATCAAAAAGGCCTGCTGAGCGCAGCTGCTCTGGCCCTCCCTGAGCTAACT	1296	
Db	444	PheGlnAlaLeuLysLysAlaLeuThrGlnAlaProValPheSerLeuPro---ThrGly	462	
Qy	1297	AAACCCCTTTACCTTTATGTGGATGAGCGTAAAGGAGTAGCCCGGGGAGTTTTAACCCAA	1356	
Db	463	GlnAspPheSerLeuTyrAlaThrGlnLysThrGlyIleAlaLeuGlyValLeuThrGln	482	
Qy	1357	ACCCTAGCACCATGAGAGACCTGTCGCTTACCTGTCAAGAGAGCTCGATCCTGTAGCC	1416	
Db	483	ValSerGlyMetSerLeuGlnProValTyrLeuSerLysGluIleAspValValAla	502	
Qy	1417	AGTGGTTGGCCCATATGCTGAAGGCTATGCGACGCTGTGGCCATCTGGTCAAGGACGCT	1476	
Db	503	LysGlyTrpProHisCysLeuTrpValMetAlaAlaValAlaValLeuValSerGluAla	522	
Qy	1477	GACAAATTCATTGGGACACAAATATAACTGTGTAATAGCCCCCATGCTATGGAGACATC	1536	
Db	523	ValLysIleIleGlnGlyArgAspLeuThrValTrpThrSerHisAspValAsnGlyIle	542	
Qy	1537	GTTCGGCAGCCCCCAGACCCGATGATGACCAACCCCGCATGACCCACTATCAAGCGCTG	1596	
Db	543	LeuThrAlaLysGlyAspLeuTrpLeuSerAspAsnHisLeuLeuAsnTyrGlnAlaLeu	562	
Qy	1597	CTTCTCCAGAGAGGGTC---ACGTTGCTCCACAGCGCTCTCAACCCCTGCCACTCTT	1653	
Db	563	LeuLeuGluGluProValLeuArgLeuArgThrCysAlaThrLeuLysProAlaThrPhe	582	
Qy	1654	CTGCTGGAAGAGACTGATGAACCACTGACTCATATTGCCATCAACTATTGATTGGAGGAG	1713	
Db	583	LeuPro---AspAsnGluGluLysIleGluHisAsnCysGlnGlnValIleAlaGlnThr	601	
Qy	1714	ACTGGGTGCGCAAGGACCTTACAGACATACCGCTGACTGGAGAAGTGTCTAACCTGGTTC	1773	
Db	602	TyrAlaAlaArgGlyAspLeuLeuGluValProLeuThrAspProAspLeuAsnLeuTyr	621	
Qy	1774	ACTCACGAGACGACTATGTGTGGAAGGTAAGAGGATGGCTGGCGGGGGTGGTGCAG	1833	
Db	622	ThrAspGlySerSerLeuAlaGluLysGlyLeuArgLysAlaGlyTyrAlaValIleSer	641	
Qy	1834	GGGACCCGACGATCTGGCGGCAGGAGCCTGCGGAAGGAACTTCAGCACAAAAGGCTGAG	1893	
Db	642	AspAsnGlyIleLeuGluSerAsnArgLeuThrProGlyThrSerAlaHisLeuAlaGlu	661	
Qy	1894	CTCATGGCCCTCAGCAAGCTTTGGCGCTGGCCGAGGGAATCCATAACATTATACG	1953	
Db	662	LeuIleAlaLeuThrTrpAlaLeuGluLeuGlyGlyLeuArgValAsnIleTyrSer	681	
Qy	1954	GACAGCAGGTATGCTTTGCGACTGCACACGTACATGGGGCATCTATAACAAAGGGGG	2013	
Db	682	AspSerLysTyrAlaTyrLeuValLeuHisAlaHisAlaIleTrpArgGluArgGlu	701	
Qy	2014	TTGCTTTACCTCAGCAGGAGGGAATAAAGAACAAAGAGGAAATCTTAAGCCTTATTGAA	2073	
Db	702	PheLeuThrSerGluGlyThrProIleAsnHisGlnGluAlaIleArgLeuLeuLeu	721	
Qy	2074	GCCGTACATTTTACCAAAAGCTAGCTATTATACACTGCTCTGCGACATCAGAAAGCTAAA	2133	
Db	722	AlaValGlnLysProLysGluValAlaValLeuHisCysGlnGlyHisGlnGluGluGlu	741	
Qy	2134	GATCTCATATCCAGAGGAAACCATAGTGCACCGGGTTGCCAGCAGCGACCCAG	2190	
Db	742	GluArgGluIleGluGlyAsnArgGlnAlaAspIleGluAlaLysLysAlaAlaArg	760	

```

1 BESEME, FREDERIC
2 BEDIN, FREDERIC
3 PARANHOS-BACCALA, GLAUCIA
4 KOMURIAN-PRADEL, FLORENCE
5 JOLIVET-REYNAUD, COLETTE
6 MANDRAND, BERNARD
7 GARSON, JEREMY
8 TURE, PHILIP
9
10 TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
11 ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR
12 THERAPEUTIC PURPOSES
13
14 NUMBER OF SEQUENCES: 210
15 CORRESPONDENCE ADDRESS:
16 ADDRESSEE: OLIFF & BERRIDGE, PLC
17 STREET: P.O. BOX 1928
18 CITY: ALEXANDRIA
19 STATE: VA
20 COUNTRY: USA
21 ZIP: 22320
22
23 COMPUTER READABLE FORM:
24 MEDIUM TYPE: Floppy disk
25 COMPUTER: IBM PC compatible
26 OPERATING SYSTEM: PC-DOS/MS-DOS
27 SOFTWARE: Patentin Release #1.0, Version #1.30
28
29 CURRENT APPLICATION DATA:
30 APPLICATION NUMBER: US/08/979,847B
31 FILING DATE: 26-No. 6582703-1997
32 CLASSIFICATION: <Unknown>
33
34 ATTORNEY/AGENT INFORMATION:
35 NAME: BERRIDGE, WILLIAM P.
36 REGISTRATION NUMBER: 30,024
37 REFERENCE/DOCKET NUMBER: WPB 39046A
38
39 TELECOMMUNICATION INFORMATION:
40 TELEPHONE: 703-836-6400
41 TELEFAX: 703-836-2787
42
43 INFORMATION FOR SEQ ID NO: 198:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 683 amino acids
46 TYPE: amino acid
47 STRANDEDNESS: single
48 TOPOLOGY: linear
49
50 MOLECULE TYPE: peptide
51
52 SEQUENCE DESCRIPTION: SEQ ID NO: 198:
53
54 US-08-979-847B-198
55
56 Alignment Scores:
57 Pred. No.: 6,19e-125 Length: 683
58 Score: 1390.50 Matches: 311
59 Percent Similarity: 59.5% Conservative: 117
60 Best Local Similarity: 43.3% Mismatches: 248
61 Query Match: 22.5% Indels: 43
62 DB: 2 Gaps: 13
63
64
65 US-10-723-552-3_COPY_2307_5741 (1-3435) x US-08-979-847B-198 (1-683)
66
67 QY 88 GTCATACCTCAGTGGCCAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATG
68 Db 1 IlemetProIuSerProThrProIuLeuGlyArgAspIleLeuAlaIysAla
69
70 QY 148 CAAATTTCTTTTGAACAAGCAAGCAAGTGTCTGCAAAATAACAAACCTATC
71 Db 21 IleIleHisLeuAsnIleGlyLys-----GlyIleProIle
72
73 QY 208 TTGACCCCTCAATTAGATGAGCAATATGCACTATATCTCTCCCTAGTAAAGCCT
74 Db 34 -----CysProLeuLeuGluGlu
75
76 QY 268 AATATACAAATTCTGGTTGGAACAGTTTCCCAAGCCTGGGCAGAAACCGCAGG
77 Db 42 Asn-----ProGluValITpAlaIleGluGlyGln
78
79 QY 328 TTGCGAAAGCAAGTTCCCCCAACAAGTTATTCAACTGAAGCGGCAGTGCACACA
80 Db 328 TTGCGAAAGCAAGTTCCCCCAACAAGTTATTCAACTGAAGCGGCAGTGCACACA

```





STREET: P.O. BOX 19928  
CITY: ALEXANDRIA  
STATE: VA  
COUNTRY: USA  
ZIP: 22320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/979,847B  
FILING DATE: 26-NO. 6582703-1997  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: BERRIDGE, WILLIAM P.  
REGISTRATION NUMBER: 30,024  
REFERENCE/DOCKET NUMBER: WPB 39046A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6400  
TELEFAX: 703-836-2787  
INFORMATION FOR SEQ ID NO: 200:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 683 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 200:  
US-08-979-847B-200  
Alignment Scores:  
Pred. No.: 6,19e-125 Length: 683  
Score: 1390.50 Matches: 311  
Percent Similarity: 59.5% Conservative: 117  
Best Local Similarity: 43.3% Mismatches: 248  
Query Match: 22.5% Indels: 43  
Gaps: 13  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-08-979-847B-200 (1-683)  
QY 88 GTCATACCTGAGTGGCCAGACCCCTCTTAGGTAGAGACTTATTCACCAAGATGGAGCA 147  
DB 1 lleMetProGluSerProThrProLeuLeuGlyArgAspIleLeuAlaLysAlaGlyAla 20  
QY 148 CAAATTTCTTTTGAACAAGGGAACAGAGTGTCTGCAAAATAACAAACCTATCACTGTG 207  
DB 21 lleIleHisLeuAsnIleGlyLys-----GlyIleProIleCys--- 33  
QY 208 TTGACCTCCCAATTAGATGACGAATATCGACTATACTCTCCCTAGTAAGCGTGATCAA 267  
DB 34 -----CysProLeuLeuGluGluGlyIle 41  
QY 268 AATATACAATTCGTGTGGAACAGTTTCCCAAGCTGGGCAGAAACCGCAGGATGGGT 327  
DB 42 Asn-----ProGluValTrpAlaIleGluGlyGlnTrpGly 53  
QY 328 TTGGCAAGACGAGTTCCCCCAAGTTATTCAACTGAAGCGCAGTGGCCACACCGATGCA 387  
DB 54 GlnAlaLysAsnAlaArgProValGlnValLysLeuLysAspSerAlaSerPheProTrp 73  
QY 388 GTCAGACAGTACCCCTTGTAGTAAAGAGCTCAAGAGGAATTCGGCCGCGATGTCCAAAGA 447  
DB 74 GlnArgLysTyProLeuArgProGluAlaLeuGlnGly\*\*GlnLysIleValLysAsp 93  
QY 448 TTAATCCAACAGGGATCCTAGTCTCTGTCCAATCTCCCTGGAATCTCCCTGTCTACCG 507  
DB 94 LeuLysAlaGlnGlyLeuValLysProCysSerSerProCysAsnThrProIleLeuGly 113  
QY 508 GTTAGAAGCGTGGGACTAATGACTATCGACACAGTACAGGACTTGAGAGAGGTCAATAAA 567  
DB 114 ValArgLysPro---AsnGlyGlnTrpArgLeuValGlnAspLeuArgIleIleAsnGlu 132

QY 568 CGGTGCGAGGATATACACCAACAGTCCCGAACCTTAACTTCTGTGTGCTCTCCCA 627  
DB 133 AlaValPheProLeuTyProAlaValSerSerProTyThrLeuLeuSerLeuIlePro 152  
QY 628 CCCCAACGGAGCTGTATACAGTATTGAGCTTAAAGGATGCTTCTTCTGTGCTGAGATTA 687  
DB 153 GluGluAlaGluTrpPheThrValLeuAspLeuLysAspAlaPhePheCysIleProVal 172  
QY 688 CACCACCTAGCCAAACACCTTTTGGCTTCGATGGAGAGATCCAGGTACGGGAGAACCC 747  
DB 173 ArgProAspSerGlnPheLeuPheAlaPheGlu-----AspPro---LeuAsnProThr 189  
QY 748 GGGCAGCTCACCTGGACCCGACTGCCCAAGGGTTCAAGAACTCCCGACCATCTTTGAC 807  
DB 190 SerGlnLeuThrTrpThrValLeuProGlnGlyPheArgAspSerProHisLeuPheGly 209  
QY 808 GAAGCCCTTACACAGAGACCTGGCCCAACTTCAGGATCCAAACACCTCAGGTGACCTCTC 867  
DB 210 GlnAlaLeuAlaGlnAspLeuSer-----GlnProSerTyLeuAspIleLeuValLeu 227  
QY 868 CAGTACGTGGATGACCTGCTTCTGCGGGAGCCACCAACAGAGACTGCTTAGAGGACG 927  
DB 228 GlnTyValAspAspLeuLeuValAlaArgSerGluThrLeuCysHisGlnAlaThr 247  
QY 928 AAGGCACCTACTGCTGGAATTGCTGACCTAGGCTACAGAGCTCTGCTAAAGAGCCGAC 987  
DB 248 GlnGluLeuLeuIlePheLeuThrThrCysGlyTyLysValSerLysProLysAlaArg 267  
QY 988 ATTTGCGAGGAGAGAGTAACATACTTGGGGTACAGTTTTCGGGACGGGACGCGATGGTG 1047  
DB 268 LeuCysSerGlnGluIleArgTyLeuGlyLeuLysLeuSerLysGlyThrArgAlaLeu 287  
QY 1048 ACGGAGGACCGAAGAAAACCTGTAGTCCAGATACCGGCCCAACACCAAGCAAAATG 1107  
DB 288 SerGluGluArgIleGlnProIleLeuAlaTyProHisProLysThrLeuLysGlnLeu 307  
QY 1108 AGAGAGTTTTTGGGACAGCTGGATTTTGCAGACTGTGTGATCCCGGGTGTTCGACCTTA 1167  
DB 308 ArgGlyPheLeuGlyIleThrGlyPheCysArgLysGlnIleProArgTyThrProIle 327  
QY 1168 GCAGCCCTCTCTACCCGCTAACCAAGAAAAGGGGAATTTCTCC----- 1212  
DB 328 AlaArgProLeuTyThrLeuIleArgGluThrGlnLysAlaAsnThrTyLeuValArg 347  
QY 1213 TGGGCTCTGAGCAGCAGAGCGATTTGATGCTATCAAAAAGGCCCTCTGAGCGCACCT 1272  
DB 348 TrpThrProThr---GluValAlaPheGlnAlaLeuLysLysAlaLeuThrGlnAlaPro 366  
QY 1273 GCTCTGGCCCTCCCTGACGTAACTAAACCTTTTACCCTTTATGTGGATGAGCGTAAGGA 1332  
DB 367 ValPheSerLeuPro---ThrGlyGlnAspPheSerLeuTyAlaThrGluLysThrGly 385  
QY 1333 GTAGCCCGGGAGTTTTTAACCCAAACCTTAGGACCATGAGAGACCTGTGCTGCTACCTG 1392  
DB 386 IleAlaLeuGlyValLeuThrGlnValSerGlyMetSerLeuGlnProValValTyLeu 405  
QY 1393 TCAAGAAGCTCGATCTCTGTAGCCAGTGGTGGCCCATATCGCTCAAGGCTATCCAGCT 1452  
DB 406 SerLysGluIleAspValValAlaLysGlyTrpProHisCysLeuTyValMetAlaAla 425  
QY 1453 GTGGCCATPACTGGTCAAGGACGCTGACAAATTTGACTTTGGGACAGAAATATACTTAATA 1512  
DB 426 ValAlaValLeuValSerGluAlaValLysIleIleGlnGlyArgAspLeuThrValTrp 445  
QY 1513 GCCCCCCATGCAATTGGAGAACATCGTTCGGCAGCCCGCCAGCCGATGATGACCAACGCC 1572  
DB 446 ThrSerHisAspValAsnGlyIleLeuThrAlaLysGlyAspLeuTyLeuSerAspAsn 465  
QY 1573 GCGATGACCCACTATCAAGGCTGCTTCTCACAGAGGGTCTC---ACGTGCTCCACCA 1629  
DB 466 HisLeuAsnTyThrGlnAlaLeuLeuGluGluProValLeuArgLeuArgThrCys 485  
QY 1630 GCGGCTCTCAACCCCTGCCACTCTTCTGCTGCAAGAGAGATGATGAACAGTGAATCATGAT 1689

```

486 AlaThrLeuGlnProAlaThrPheLeuPro---AspAenGluGluGlnIleGluHisAen 504
1690 TGCCATCAACTATTGATGAGGAGACATGGGGTCCGCAAGGACCTTACAGACATACCGCTG 1749
505 CysGlnGlnValIleAlaGlnThrTyrAlaAlaArgGlyAspLeuLeuGluValProLeu 524
1750 ACTGGAGAGTGCTAACTGCTGCTACGACGAGAGCTATGTTGGTGAAGTAAAGAGG 1809
525 ThrAspProAspLeuAenLeuTyrThrAspGlySerLeuAlaGluLysGlyLeuArg 544
1810 ATGCTCGGGGGGGTGGTGGAGCGGACCGCAGCATCTGGGCGCAGCAGCTTCCCGGAA 1869
545 LysAlaGlyTyrAlaValIleSerAspAenGlyIleLeuGluSerAsnArgLeuThrPro 564
1870 GGAATCTTCAGCAACAAAGGCTGAGCTCATGGCCCTTCACGCAAGCTTTCGGCTGGCGAA 1929
565 GlyThrSerAlaHisLeuAlaGluLeuIleAlaLeuThrTrpAlaLeuGluLeuGlu 584
1930 GGGAAATCCATAAATTTATACGACAGAGGATGCTTGGCTTGGCTGACACATACAT 1989
585 GlyLysArgValAsnIleTyrSerAspSerLysTyrAlaTyrLeuValLeuHisAlaHis 604
1990 GGGGCCCATATAACAAGGGGCTGCTTACCTCAGCAGGAGGGAATAAAGAAACAAA 2049
605 AlaAlaIleTrpArgGluArgGluPheLeuThrSerGluGlyThrProIleAsnHisGln 624
2050 GAGAAATCTTAAGCTTATTAGAGCGGTACATTTACCAAAAGGCTAGCTATTATATACAC 2109
625 GluAlaIleArgLeuLeuLeuAlaValGlnLysProLysGluValAlaValLeuHis 644
2110 TGTCTCGACATCAGAAAGCTAAAGATCTCATATCCAGAGAAACACGATGGCTGACCGG 2169
645 CysGlnGlyHisGlnGluGluGluArgGluIleGluGlyAsnArgGlnAlaAspIle 664
2170 GTTGCCAAAGCAGGAGCC---CAGGGTGTTAACTTCTGCTATATAGAAATGCC 2223
665 GluAlaLysLysAlaAlaArgGlnAspSerProLeuGluMetLeuIleGluGlyPro 683

RESULT 11
US-08-979-847B-208
; Sequence 208, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847B
; FILING DATE: 26-No. 6582703-1997
; CLASSIFICATION: <Unknown>

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-8400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 683 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-08-979-847B-208

Alignment Scores:
Pred. No.: 1,21e-124 Length: 683
Score: 1387.50 Matches: 311
Percent Similarity: 59.4% Conservative: 116
Best Local Similarity: 43.3% Mismatches: 249
Query Match: 22.4% Indels: 43
DB: 2 Gaps: 13

US-10-723-552-3_COPY_2307_5741 (1-3435) x US-08-979-847B-208 (1-683)

QY 88 GTCATCTCAGTGGCCAGCACCGCTCTTAGGTAGAGACTTATTGACCAAGATGGGAGCA 147
Db 1 IleMetProGluSerProThrProLeuLeuGlyArgAspIleLeuAlaGlyAla 20
QY 148 CAAATTTCTTTTGAACAAGGGAACCAAGAGTGTCTGCAATAACAAACCTATCCTGTG 207
Db 21 IleIleHisLeuAsnIleGlyLys-----GlyIleProIleCys--- 33
QY 208 TTGACCTCCAAATTAGATGAGAAATATCGACTATCTCTCCCTAGTAAAGCTGATCAA 267
Db 34 -----CysProLeuLeuGluGluGlyIle 41
QY 268 AATATACAATTCTGTGGAAACAGTTTCCCAAGCCCTGGGCGAGAAACCGCAGGATGGGT 327
Db 42 Asn-----ProGluValTrpAlaIleGluGlyGlnTyrGly 53
QY 328 TTGGCAAGCAAGTTCCCCCAAGTTATTCAACTGAAGCCAGTGCACACACCATGTGTCA 387
Db 54 GlnAlaLysAsnAlaArgProValGlnValLysLeuLysAspSerAlaSerPheProTyr 73
QY 388 GTCAGACAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTCGGCCGATGTCCAAAGA 447
Db 74 GlnArgLysTyrProLeuArgProGluAlaLeuGlnGly**GlnLysIleValLysAsp 93
QY 448 TTAATCCACAGGGCATCTCTAGTTCTCTGTCCTCAATCTCCCTGGAATCTCCCTGCTACCG 507
Db 94 LeuLysAlaGlnGlyLeuValLysProCysSerSerProCysAsnThrProIleLeuGly 113
QY 508 GTTAGAAGCCTGGGACTAATGACTATCGACACAGTACAGGACTTTGAGAGAGGTCAATAAA 567
Db 114 ValArgLysPro---AsnGlyGlnTrpArgLeuValGlnAspLeuArgIleIleAsnGlu 132
QY 568 CGGGTGCAGGATATACACCAACAGTCCCGAACCCTTTATACCTCTTGTGTCTCTCCCA 627
Db 133 AlaValPheProLeuTyrProAlaValSerSerProTyrThrLeuLeuSerLeuIlePro 152
QY 628 CCCCAACGGAGCTGTATACAGTATTGGACTTAAAGGATGCCTTCTCTGCTGAGATTA 687
Db 153 GluGluAlaGluTrpPheThrValLeuAspLeuLysAspAlaPhePheCysIleProVal 172
QY 688 CACCCCACTAGCCCAACACACTTTTTCGCTTCGAAATGGAGAGATCCAGTACGGGAAGAAC 747
Db 173 ArgProAspSerGlnPheLeuPheAlaPheGlu-----AspPro---LeuAsnProThr 189
QY 748 GGCAGCTCACCTGGACCCGACTGCCCAAGGGTTCACAAAGGTTCAAGAACTCCCGCACCATCTTTGAC 807
Db -----

```

Db 190 SerGlnLeuThrTrpThrValLeuProGlnGlyPheArgAspSerProHisLeuPheGly 209  
QY 808 GAAGCCCTACAGAGAGCTGCCAACTTCAGGATCCAAACCCCTCAGGTGACCTCCTC 867  
Db 210 GlnAlaLeuAlaGlnAspLeuSer-----GlnProSerTyrLeuAspThrLeuValLeu 227  
QY 868 CAGTACGTGGATGACCTCTCTGGCGGAGCCCAACAGGAGCTCTAGAAAGGCACG 927  
Db 228 GlnTyrValAspLeuLeuValAlaArgSerGlnThrLeuCysHisGlnAlaThr 247  
QY 928 AAGGCACTACTGCTGGAATTGCTGACCTAGGCTACAGAGCTCTGCTAAAGAGCCGAC 987  
Db 248 GlnGluLeuLeuIlePheLeuThrThrCysGlyTyrLysValSerLysProLysAlaArg 267  
QY 988 ATTTGACGAGAGAGTAACACTATTGGGGTACAGTTTGGGGGACGGCAGCAGTGGCTG 1047  
Db 268 LeuCysSerGlnGluIleArgTyrLeuGlyLeuLysLeuSerLysGlyThrArgAlaLeu 287  
QY 1048 ACGGAGGCACGGAAGAACTGTAGTCCAGATACCGGCCCAACACACAGCCCAACAATG 1107  
Db 288 SerGluGluArgIleGlnProIleLeuAlaTyrProHisProLysThrLeuLysGlnLeu 307  
QY 1108 AGAGAGTTTTGGGACAGCTGGATTGTCAGACTGTGGATCCCGGGTGTTCGACACTTA 1167  
Db 308 ArgGlyPheLeuGlyIleThrGlyPheCysArgLysGlnIleProArgTyrThrProile 327  
QY 1168 GCAGCCCACTCTACCCCTAACCAAGAAAGGGGAATTCCTCC----- 1212  
Db 328 AlaArgProLeuTyrThrLeuIleArgGluThrGlnLysAlaAsnThrTyrLeuValArg 347  
QY 1213 TGGGCTCTGAGCACCAAGGCATTTGATGCTATCAAAAGGCCCTGCTGAGCGCACCT 1272  
Db 348 TrpThrProThr---GluValAlaPheGlnAlaLeuLysLysAlaLeuThrGlnAlaPro 366  
QY 1273 GCTCTGGCCCTCCCTGACGTAACTAAACCTTTTACCTTTATGTGGATGAGCGTAAGGGA 1332  
Db 367 ValPheSerLeuPro---ThrGlyGlnAspPheSerLeuTyrAlaThrGluLysThrGly 385  
QY 1333 GTAGCCCGGGAGATTTAACCAACCTTAGCACCATGAGACCATGAGAACCTGTGCTACCTG 1392  
Db 386 IleAlaLeuGlyValLeuThrGlnValSerGlyMetSerLeuGlnProValValTyrLeu 405  
QY 1393 TCAAGAAAGCTCGATTCGTGACCGAGTGTGGCCCATATGCTGAAGGCTATGCGACGT 1452  
Db 406 SerLysGluIleAspValValAlaLysGlyTrpProHisCysLeuTyrValMetAlaAla 425  
QY 1453 GTGGCCATCTGCTCAAGGACGCTGACAAATGACTTTGGGACAGAAATATAACTGTATA 1512  
Db 426 ValAlaValLeuValSerGluAlaValLysIleIleGlnGlyArgAspLeuThrValTrp 445  
QY 1513 GCCCCCATCATTTGGAGAAATGCTTTCGGCAGCCCGCCAGCCGATGATGACCAACGCC 1572  
Db 446 ThrSerHisAspValAsnGlyIleLeuThrAlaLysGlyAspLeuTrpLeuSerAspAsn 465  
QY 1573 CGCATGCCACTATCAAGCTGCTTCTCACAGAGGGTCT---ACGTTGCTCCACCA 1629  
Db 466 HisLeuLeuAsnTyrGlnAlaLeuLeuLeuGluProValLeuArgLeuArgThrCys 485  
QY 1630 GCGGCTCTCAACCTCGCCACTCTTCTGCTGAAAGAGACTGTAGTACCACTGACTCATGAT 1689  
Db 486 AlaThrLeuGlnProAlaThrPheLeuPro---AspAsnGluGluLysIleGluHisAsn 504  
QY 1690 TGCCATCAACTATTGATTGAGAGACTGGGGTCCGAAGGACCTTACAGACATACCGCTG 1749  
Db 505 CysGlnGlnValIleAlaGlnThrTyrTrpAlaAlaArgGlyAspLeuLeuGluValProLeu 524  
QY 1750 ACTGGAGAGTGTCAACCTGTTCTACTGACGAACGACTATGTGTGGAAGGTAAAGG 1809  
Db 525 ThrAspProAspLeuAsnLeuTyrThrAspGlySerSerLeuAlaGluLysGlyLeuArg 544  
QY 1810 ATGGCTGGGGCGGTGGTACCGGGACCCGACGATCTGGCCGACGAGCTGCGCGAA 1869  
Db 545 LysAlaGlyTyrAlaValIleSerAspAsnGlyIleLeuGluSerAsnArgLeuThrPro 564

QY 1870 GGAACCTTCAGCAAAAGGCTGAGCTCATGSCCTCAGCAAGCTTTTGGCGCTGCGCGAA 1929  
Db 565 GlyThrSerAlaHisLeuAlaGluLeuIleAlaLeuThrTrpAlaLeuGluLeuGlyGlu 584  
QY 1930 GGGAAATCCATAAATATATATACGACAGCAGGATGCTTTTGGGACTGCGACACGTACAT 1989  
Db 585 GlyLysArgValaAnIleTyrSerAspSerLysTyrAlaTyrLeuValLeuHisAlaHis 604  
QY 1990 GGGGCCATCTATAAACAAGGGGTTGCTTACCTCAGCAGGAGGAGGAAATAAGAACAA 2049  
Db 605 AlaAlaIleTrpArgGluArgPheLeuThrSerGluGlyThrProIleAsnHisGln 624  
QY 2050 GAGGAATCTTAACCTATTAGAACCGCTACATTTACCAAAAAGGCTAGCTATTATACAC 2109  
Db 625 GluAlaIleArgArgLeuLeuAlaValGlnLysProLysGluValAlaValLeuHis 644  
QY 2110 TGTCTCGACATCAGAAAGCTTAAAGATCTCATATCCAGAGAAACACAGATGGCTGACCGG 2169  
Db 645 CysGlnGlyHisGlnGluGluGluArgGluLeuGluArgGlnAlaAspIle 664  
QY 2170 GTTCCCAAGCAGCGAGCC---CAGGGTGTAACTTCTGCTCTATAATAAGAAATGCC 2223  
Db 665 GluAlaLysLysAlaAlaArgGlnAspSerProLeuGluMetLeuIleGluGlyPro 683

## RESULT 12

US-08-979-847B-210  
; Sequence 210, Application US/08979847B  
; Patent No. 6582703  
; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE  
; BESEME, FREDERIC  
; BEDIN, FREDERIC  
; PARANHOS-BACCALA, GLAUCIA  
; KOMURIAN-PRADEL, FLORENCE  
; JOLIVET-REYNAUD, COLETTE  
; MANDRAND, BERNARD  
; GARSON, JEREMY  
; TUKE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL  
; THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC

STREET: P.O. BOX 19928

CITY: ALEXANDRIA

STATE: VA

COUNTRY: USA

ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,847B

FILING DATE: 26-No. 6582703-1997

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 39046A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 210:

SEQUENCE CHARACTERISTICS:

LENGTH: 683 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

## SEQUENCE DESCRIPTION: SEQ ID NO: 210:

US-08-979-847B-210

## Alignment Scores:

Pred. No.:	1,21e-124	Length:	683
Score:	1387.50	Matches:	311
Percent Similarity:	59.4%	Conservative:	116
Best Local Similarity:	43.3%	Mismatches:	249
Query Match:	22.4%	Indels:	43
DB:	2	Gaps:	13

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-08-979-847B-210 (1-683)

```
Qy 88 GTCATACCTGAGTCCCGACAGCCCTCTTAGGTAGAGACTTATTGACCAAGATGGAGCA 147
Dy 1 IleMetProGluSerProThrProLeuLeuGlyArgAspIleLeuAlaLysAlaGlyAla 20
Qy 148 CAATTTCTTTTGAACAGGGAACACAGAGTGTCTGCAATATACAAACCTATCACTGTG 207
Dy 21 IleIleHisLeuAsnIleGlyLys-----GlyIleProIleCys--- 33
Qy 208 TTGACCCCTCCAATTAGATGACGAATATCGATATACTCTCCCTAGTAAAGCCTGATCAA 267
Dy 34 -----CysProLeuLeuGluGluGlyIle 41
Qy 268 AATATACAAATTCGTTGGAACAGTTTCCCAAGCCTCGGCAGAAACCGCAGGGATGGT 327
Dy 42 Asn-----ProGluValTrpAlaIleGluGlyGlnTrpGly 53
Qy 328 TTGCAAGCAAGTTCCTCCCAACAGTTATTCACTGAAGCCAGTGCACACAGTGTCA 387
Dy 54 GlnAlaLysAsnAlaArgProValGlnValLysLeuLysAspSerAlaSerPheProTrp 73
Qy 388 GTACAGACAGTACCCCTTGAGTAAAGAGCTCAAGAAGAAATTCGCGCGCATGTCCAAGA 447
Dy 74 GlnArgLysTrpProLeuArgProGluAlaLeuGlnGly***GlnLysIleValLysAsp 93
Qy 448 TTAATCAACAGGGCATCTCTAGTTCCTGTCCTCAATCTCCCTGGAATCTCCCTGCTACCG 507
Dy 94 LeuLysAlaGlnGlyLeuValLysProCysSerSerProCysAsnThrProIleLeuGly 113
Qy 508 GTTAGAAGCCTGGGACTAATGACTATCGACAGTACAGGCTTGAGAGAGTCAATAAA 567
Dy 114 ValArgLysPro---AsnGlyGlnTrpArgLeuValGlnAspLeuArgIleIleAsnGlu 132
Qy 568 CGGTCGAGGATATACACCAACAGTCCCGAACCTTATACCTCTTGTGTGCTCTCCCA 627
Dy 133 AlaValPheProLeuTrpProAlaValSerSerProTrpTrpTrpLeuLeuSerLeuIlePro 152
Qy 628 CCCCAACGGAGCTGGTATACAGTATTGGACTTAAAGGATGCTTCTTCTGCTGAGATTA 687
Dy 153 GluGluAlaGluTrpPheTrpValLeuAspLeuLysAspAlaPhePheCysIleProVal 172
Qy 688 CACCCCACTAGCCAAACCACTTTTTCCTTGAATGGAGAGATCCAGTACGGGAAGAAC 747
Dy 173 ArgProAspSerGlnPheLeuPheAlaPheGlu-----AspPro---LeuAsnProThr 189
Qy 748 GGGCAGCTCACCTGACCGACTGCCCCAAGGGTTCAAGAACTCCCGACCATCTTTGAC 807
Dy 190 SerGlnLeuThrTrpTrpValLeuProGlnGlyPheArgAspSerProHisLeuPheGly 209
Qy 808 GAAGCCCTCACACAGAGACTGCGCAACTTCAGGATCCAAACCCCTCAGGTGACCCCTCCTC 867
Dy 210 GlnAlaLeuAlaGlnAspLeuSer-----GlnProSerTrpLeuAspThrLeuValLeu 227
Qy 868 CAGTACCTGATGACTGCTTCTGCGGGAGCCACCAACAGGACTGCTTAGAAGGCACG 927
Dy 228 GlnTrpValAspAspLeuLeuValAlaArgSerGlnTrpLeuCysHisGlnAlaThr 247
Qy 928 AAGGCACCTACTGCTGGAATGTCTGACCTAGGCTACAGAGCCTCTGCTAGAAGGCCACG 987
Dy 248 GlnGluLeuLeuIlePheLeuThrTrpCysGlyTrpLysValSerLysProLysAlaArg 267
```

```
Qy 988 ATTTCAGGACAGAGAGGTAAACATACTTGGGTACAGTTTGGGGACGGCAGCGATGGCTG 1047
Dy 268 LeuCysSerGlnGluIleArgTrpLeuGlyLeuLysLeuSerLysGlyThrArgAlaLeu 287
Qy 1048 ACGGAGCAGCGAAGAAATCTGTAGTCCAGTACCGGCCCAACACAGCCAAACAAATG 1107
Dy 288 SerGluGluArgGlnGlnProIleLeuAlaTrpProHisProLysTrpTrpLeuLysGlnLeu 307
Qy 1108 AGACAGTTTTTGGGGACAGCTGGATTTCGAGACTGTGGATCCCGGGTTTTCGACCTT 1167
Dy 308 ArgGlyPheLeuGlyIleThrGlyPheCysArgLysGlnIleProArgTrpTrpProIle 327
Qy 1168 GCAGCCCACTCTACCCGCTAAACCAAGAAAGGGAATTCCTC----- 1212
Dy 328 AlaArgProLeuTrpTrpLeuIleArgGluThrGlnLysAlaAsnThrTrpLeuValArg 347
Qy 1213 TGGGCTCTCAGACACAGAGGCAATTTGATGCTATCAAAAGGCCCTCTGCTGAGCGCACT 1272
Dy 348 TrpThrProThr---GluValAlaPheGlnAlaLeuLysAlaLeuThrGlnAlaPro 366
Qy 1273 GCTCTGCCCTCCTCGACGTAAACCTTTTACCTTTTATGTTGGATGAGCGTAAGGA 1332
Dy 367 ValPheSerLeuPro---ThrGlyGlnAspPheSerLeuTrpAlaThrGluLysThrGly 385
Qy 1333 GTAGCCCGGGAGTTTAAACCAACCTAGGACCATGGAGAACCTGTGCTGCTACCTG 1392
Dy 386 IleAlaLeuGlyValLeuThrGlnValSerGlyMetSerLeuGlnProValValTrpLeu 405
Qy 1393 TCAAGGAAGCTCGATCTGTAGCAGTGTGGCCCATATGCTTGAAGGCTATCGCAGCT 1452
Dy 406 SerLysGlnIleAspValValAlaLysGlyTrpProHisCysLeuTrpValMetAlaAla 425
Qy 1453 GTGCCCATACTGTGTAAGGACCGCTGACAAATTTGACTTTGGGACAGAAATAAAGTGAATA 1512
Dy 426 ValAlaValLeuValSerGluAlaValLysIleIleGlnGlyArgAspLeuThrValTrp 445
Qy 1513 GCCCCCCCATGATGGAGAACATCGTTTCGGAGCCCCCAGACCATGATGATGACCAAGCC 1572
Dy 446 ThrSerHisAspValAsnGlyIleLeuThrAlaLysGlyAspLeuTrpLeuSerAspAsn 465
Qy 1573 CGCATGACCACTATCAAGCCTGCTTCTCACAGAGAGGCTC---ACGTTGCTCCACCA 1629
Dy 466 HisLeuLeuAsnTrpGlnAlaLeuLeuGluGluProValLeuArgLeuArgThrCys 485
Qy 1630 GCGCTCTCAACCTGCTGCTTCTGCTGGAAGAGACTGATGAACAGTGAAGTCAATGAT 1689
Dy 486 AlaThrLeuGlnProAlaThrPheLeuPro---AspAsnGluGluLysIleGluHisAsn 504
Qy 1690 TGGCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTG 1749
Dy 505 CysGlnGlnValIleAlaGlnThrTrpAlaAlaArgGlyAspLeuLeuGluValProLeu 524
Qy 1750 ACTGGAGAGTGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1809
Dy 525 ThrAspProAspLeuAsnLeuTrpAspGlySerSerLeuAlaGluLysGlyLeuArg 544
Qy 1810 ATGCTGGGGCGCGCTGGTGGAGCGGACCCGACAGCTCTGGGCGACAGCTCCCGCGAA 1869
Dy 545 LysAlaGlyTrpAlaValIleSerAspAsnGlyIleLeuGluSerAsnArgLeuThrPro 564
Qy 1870 GGAACCTTCAGCAAAAGGCTGAGCTCATGCGCCCTCACGCAAGCTTTGCGGCTGCCCAA 1929
Dy 565 GlyThrSerAlaHisLeuAlaGluLeuIleAlaLeuThrTrpAlaLeuGluGlyGlu 584
Qy 1930 GGGAAATCCATAAATTTATACGACAGCAGTATGCTTTCGCTGCTGCTGCTGCTGCTGCT 1989
Dy 585 GlyLysArgValAsnIleTrpSerAspSerLysTrpAlaTrpLeuValLeuHisAlaHis 604
Qy 1990 GGGGCCATCTATAAACAAGGGGTTGCTTACCTCAGCAGGAGGGGAAATAAAGAACAAA 2049
Dy 605 AlaAlaIleTrpArgGluArgGluPheLeuThrSerGluGlyThrProIleAsnHisGln 624
Qy 2050 GAGGAAATCTAAGCCTATTAGAACCGGTACATTTTACCAAAAAAGGCTAGCTATTATACAC 2109
```

Db 625 GluAlaIleArgLeuLeuLeuAlaValGlnLysProLysGluValAlaValLeuHis 644  
Qy 2110 TGTCCTGACATCAGAAAGCTAAGATCTCATATCCAGAGAAACAGATGGCTGACCGG 2169  
Db 645 CysGlnGlyHisGlnGluGluArgGluLeuGluGluArgGlnAlaAspIle 664  
Qy 2170 GTTCCAGCAGGAGGAGCC--CAGGGTCTTAACCTTCTGCCTATATAGAAATGCC 2223  
Db 665 GluAlaLysLysAlaAlaArgGlnAspSerProLeuGluMetLeuLeuGluGlyPro 683

RESULT 13  
US-08-979-847B-91  
; Sequence 91, Application US/08979847B  
; Patent No. 6582703  
; GENERAL INFORMATION:  
; APPLICANT: PERRON, HERVE  
; BESEME, FREDERIC  
; BEDIN, FREDERIC  
; PARANHOS-BACCALA, GLAUCIA  
; KOMURIAN-FRADEL, FLORENCE  
; JOLIVET-REYNAUD, COLETTE  
; MANDRAND, BERNARD  
; GARSON, JEREMY  
; TUKE, PHILIP  
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL  
; THERAPEUTIC PURPOSES  
; NUMBER OF SEQUENCES: 210  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OLIFF & BERRIDGE, PLC  
; STREET: P.O. BOX 19928  
; CITY: ALEXANDRIA  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/979,847B  
; FILING DATE: 26-NO. 6582703-1997  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERRIDGE, WILLIAM P.  
; REGISTRATION NUMBER: 30,024  
; REFERENCE/DOCKET NUMBER: WPB 39046A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-836-6400  
; TELEFAX: 703-836-2787  
; INFORMATION FOR SEQ ID NO: 91:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 654 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 91:  
US-08-979-847B-91

Alignment Scores:  
Pred. No.: 3,02e-119 Length: 654  
Score: 1331.50 Matches: 286  
Percent Similarity: 61.9% Conservative: 109  
Best Local Similarity: 44.8% Mismatches: 228  
Query Match: 21.5% Indels: 15  
DB: 2 Gaps: 9

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-08-979-847B-91 (1-654)  
Qy 295 CCCCAAGCTTGGCAGAAACCGCAGGATGGTTTGGCAAGCAAGTTCCCCCACAAGTT 354

Db 18 ProGluValIrpAlaIleGluGlyGlnTyrGlyGlnAlaLysAsnAlaAArgProValGln 37  
Qy 355 ATTCAACTGAAGCCAGTCGCCACACAGTGTCACTAGACAGTACCCCTTGAGTAAGAA 414  
Db 38 ValLysLeuLysAspSerAlaSerPheProTyrGlnArgLysTyrProLeuArgProGlu 57  
Qy 415 GCTCAAGAAGAAATTCGGCCGATGTCACAAAGATTAATCAACAGGCGATCCTAGTTCT 474  
Db 58 AlaLeuGlnGly\*\*GlnLysIleValLysAspLeuLysAlaGlnGlyLeuValLysPro 77  
Qy 475 GTCCAATCTCCCTGGAATACTCCCTGCTACCCGGTTAGAAAGCCTGGGACTAATGACTAT 534  
Db 78 CysSerSerProCysAsnThrProIleLeuGlyValArgLysPro---AsnGlyGlnIrp 96  
Qy 535 CGACCACTACAGACTTGAAGAGGTCAATAAACGGGTGCGAGTATATACACCCACAGTC 594  
Db 97 ArgLeuValGlnAspLeuArgIleIleAsnGlnAlaValPheProLeuTyrProAlaVal 116  
Qy 595 CCGAACCTTATAACCTCTGTGTCTCCACCCCAACGGAGCTGTGTATACAGTATTG 654  
Db 117 SerSerProTyrThrLeuLeuSerLeuIleProGluGlnAlaGluIrpPheThrValLeu 136  
Qy 655 GACTTAAAGGATGCTTCTTCTGCTGAGATTACACCCCACTAGCCCAACCACTTTTGTCC 714  
Db 137 AspLeuLysAspAlaPhePheCysIleProValArgProAspSerGlnPheLeuPheAla 156  
Qy 715 TTCGAATGGAGATCCAGGTACGGGAAGAACCGGCAGCTCAGCTGAGCCGAGTCCGCC 774  
Db 157 PheGlu-----AspPro---LeuAsnProThrSerGlnLeuThrIrpThrValLeuPro 173  
Qy 775 CAAGGGTTCAGAACTCCCGACCATCTTTGACCAAGCCCTACACAGAGACTGGCCCAAC 834  
Db 174 GlnGlyPheArgAspSerProHisLeuPheGlyGlnAlaLeuAlaGlnAspLeuSerGln 193  
Qy 835 TTCAGGATCCAAACCTCCAGGTACCCCTCTCCAGTACGTGGATGAGCTCTCTTCTGCG 894  
Db 194 Phe-----SerTyrLeuAspThrLeuValLeuGlnTyrValAspAspLeuLeuVal 211  
Qy 895 GGACCCACCAACAGGACTGCTTAGAGGACGACGAGGCACTACTGCTGGAATTTGCTGAC 954  
Db 212 AlaArgSerGluThrLeuCysHisGlnAlaThrGlnGluLeuLeuThrPheLeuThr 231  
Qy 955 CTAGCTACAGAGCTCTGCTAAGAAGCCAGATTTCAGAGAGAGGAGTAACTACTGTG 1014  
Db 232 CysGlyTyrLysValSerLysProLysAlaArgLeuCysSerGlnGluIleArgTyrLeu 251  
Qy 1015 GGGTACAGTTTTCGGGACGGGACGATGGCTGACGGAGGACCGAAGAAACCTGTAGTC 1074  
Db 252 GlyLeuLysLeuSerLysGlyThrArgAlaLeuSerGluGluArgIleGlnProIleLeu 271  
Qy 1075 CAGATACGGGCCCCAACCAACAGCCCAACAAATGAGAGAGTTTTCGGGACAGCTGGATT 1134  
Db 272 AlaTyrProHisProLysThrLeuLysGlnLeuArgGlyPheLeuGlyIleThrGlyPhe 291  
Qy 1135 TGCAGACTGTGGATCCCGGGGTTTCGACCTTAGCAGCCCACTCTACCCGCTAACCAAA 1194  
Db 292 CysArgLysGlnIleProArgTyrThrProIleAlaArgProLeuTyrThrLeuIleArg 311  
Qy 1195 GAAAAAGGGGAATTCCTCC-----TGGGCTCTGAGCACACCAAGGACATTT 1239  
Db 312 GluThrGlnLysAlaAsnThrTyrLeuValArgIrpThrProThr---GluValAlaPhe 330  
Qy 1240 GATGCTATCAAAAAGGCCCTCTGAGCGCAGCTGCTGCGCCCTCCCTGAGCTAACTAAA 1299  
Db 331 GlnAlaLeuLysLysAlaLeuThrGlnAlaProValPheSerLeuPro---ThrGlyGln 349  
Qy 1300 CCCTTTACCTTTATGTGATGAGGTAAAGGAGTAGCCCGGGAGTTTAAACCCAAACC 1359  
Db 350 AspPheSerLeuTyrAlaThrGluLysThrGlyIleAlaLeuGlyValLeuThrGlnVal 369  
Qy 1360 CTAGACCATCGAGAGACCTGTGCGCTTACCTGTCAAGAAAGCTCGATTCGTGTAGCCAGT 1419

```
Db 370 SerGlyMetSerLeuGlnProValValTyrLeuSerLysGluIleAspValValAlaLys 389
Qy 1420 GTTGGCCCATATGCTGGAAGCTATCGCAGCTGTGGCCATCTGTCAGGAGCGCTGAC 1479
Db 390 GlyTrpProHisCysLeuTyrValMetAlaValAlaValLeuValSerGluAlaVal 409
Qy 1480 AAATGTGACTTTGGGACAAATATACTGTAATAGCCCCCATGATGGAGAAATCGTT 1539
Db 410 LysIleIleGlnGlyArgAspLeuThrValTyrThrSerHisAspValAsnGlyIleLeu 429
Qy 1540 CGGCAGCCCCCAGCCGATGACCAACGCCCGCATGACCCACTATCAAGAGCTGTT 1599
Db 430 ThrAlaLysGlyAspLeuTyrPheSerAspHisLeuAsnTyrGlnAlaLeuLeu 449
Qy 1600 CTCACAGAGAGGTC---ACGTGCTCCACAGCGCTCTCAACCCCTGCCACTCTTCG 1656
Db 450 LeuGluGluProValLeuArgLeuThrCysAlaThrLeuLysProAlaThrPheLeu 469
Qy 1657 CCTGAGAGAGTGTGAACACGAGTCACTATGTCCTCACTGATGTCATGTTGATTGAGAGACT 1716
Db 470 Pro---AspAsnGluGluLysIleGluHisAsnCysGlnGlnValIleAlaGlnThrTyr 488
Qy 1717 GGGGTCCGACAGGACCTTACAGACATACCGCTGACTGGAGAGTGTAACTGGTTCCT 1776
Db 489 AlaAlaArgGlyAspLeuLeuGluValProLeuThrAspProAspLeuAsnLeuTyrThr 508
Qy 1777 GACGGAAGCAGCTATGTGGTGAAGTAAGAGGATGGCTGGCGCGGGTGGTGGACGGG 1836
Db 509 AspGlySerSerLeuAlaGluLysGlyLeuArgLysAlaGlyTyrAlaValIleSerAsp 528
Qy 1837 ACCGCGACGATCTGGCGAGCAGCTGCGGGAAGAACTTCAGCACAAAAGCTGAGCTC 1896
Db 529 AsnGlyIleLeuGluSerAsnArgLeuThrProGlyThrSerAlaHisLeuAlaGluLeu 548
Qy 1897 ATGCGCTTCAGCAAGCTTTCGGCTGCGCCAGGGAATCCATAACATTTATACGGAC 1956
Db 549 IleAlaLeuThrTrpAlaLeuGluLeuGlyGlyGlyLysArgValAsnIleTyrSerAsp 568
Qy 1957 AGCAGGTATGCTTTGGGACTGTCACAGCTATACGCGGCTCTATAACAAAGGGGTG 2016
Db 569 SerLysTyrAlaTyrLeuValLeuHisAlaHisAlaIleTyrArgGluArgGluPhe 588
Qy 2017 CTTACCTCAGCAGGAGGGAATTAAGAACAAAGAGGAATCTTAAGCTTATTAGAAGCC 2076
Db 589 LeuThrSerGluGlyThrProIleAsnHisGlnGluAlaIleArgLeuLeuAla 608
Qy 2077 GTACATTTACCAAAAGCTAGCTATTATACACTGCTCTGACATCAGAAGCTAAAGAT 2136
Db 609 ValGlnLysProLysGluValAlaValIleuHisCysGlnGlyHisGlnGluGluGlu 628
Qy 2137 CTCATATCCAGAGAAACACGATGGCTGACCGGGTTCGCAAGCAGGAGGCCAG 2190
Db 629 ArgGluIleGluGlyAsnArgGlnAlaAspIleGluAlaLysLysAlaAlaArg 646
```

## RESULT 14

```
US-09-120-653D-5
; Sequence 5, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120,653D
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: KopatentIn 1.71
```

```
; SEQ ID NO 5
; LENGTH: 768
; TYPE: PRT
; ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
; FEATURE:
; NAME/KEY: NON_CONS
; LOCATION: (46)
; OTHER INFORMATION: any one of amino acids
; FEATURE:
; NAME/KEY: NON_CONS
; LOCATION: (193)
; OTHER INFORMATION: any one of amino acids
; FEATURE:
; NAME/KEY: NON_CONS
; LOCATION: (322)
; OTHER INFORMATION: any one of amino acids
; FEATURE:
; NAME/KEY: NON_CONS
; LOCATION: (374)
; OTHER INFORMATION: any one of amino acids
; FEATURE:
; NAME/KEY: NON_CONS
; LOCATION: (526)
; OTHER INFORMATION: any one of amino acids
; FEATURE:
; NAME/KEY: NON_CONS
; LOCATION: (586)
; OTHER INFORMATION: any one of amino acids
US-09-120-653D-5
```

## Alignment Scores:

Pred. No.:	5.65e-110	Length:	768
Score:	1236.00	Matches:	305
Percent Similarity:	55.2%	Conservative:	106
Best Local Similarity:	41.0%	Mismatches:	270
Query Match:	20.0%	Indels:	63
DB:	2	Gaps:	18

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-09-120-653D-5 (1-768)

```
Qy 73 ACCACCTGTTTCTGTCATACCTGAGTCCGACGACCCCTCTTAGGTAGAGACTTATTG 132
Db 80 SerHisAlaPheLeuValIleProGluSerProThrProLeuLeuGlyArgAspIleLeu 99
Qy 133 ACCAAGATGGAGCACAAATTTCTTTTGAACAGGAAACCCAGAAAGTGTCTGCATAAATAC 192
Db 100 AlaLysAlaGlyAlaIleThr-TyrMetAsnLysArgAsnLys----- 113
Qy 193 AAACCTATCACTGTGTGACCCCTCCATATAGATGACGAATATCGACTATATCTCTCCCTA 252
Db 114 ----LeuThrIleCysCysPro-----LeuLeuGluArgGluSerThrLeu---- 127
Qy 253 GTAAGCCTGATCAAAATATACAAATCTGTTGGAACAGTTTCCCCAAGCCTGGGCGAGAA 312
Db 128 -----LysSerGlyHisTrpLysAspAsnLeuGluGlyGlnLy 140
Qy 313 ACCGAGGGATGGTTTGGCAAGAGTTCCTCCCAAGTTATTCACTGAAGGCCAGT 372
Db 140 sMetProAlaGlnSerLysSerGlyLysArgProHisPhe----- 154
Qy 373 GCCACACCACTGTCAGTCAGACAGTACCCCTTGGATTAAGAAAGCTCAAGAGGAATTCGG 432
Db 155 -----SerTyrGln--ArgGlnTyrProLeuArgProGluAlaHisLysGlyLeuGln 171
Qy 433 CCGCATGTCCAAAGATTAATTCACAGGGGATCTCTAGTTCTCTGTCCTTCCTTCGGAAT 492
Db 172 AspIleAlaGluHisLeuIleAlaGlnGlyLeuValArgLysCysSerSerPro---Asn 190
Qy 493 ACTCCC---CTGCTACCGGTTAGAAAGCCTGGGAGTAATGACTATCGACAGGTACAGGAC 549
Db 191 ThrPro***IleLeuGlyValGlnLysPro---AsnGlyGlnTyrArgLeuValGlnAsp 209
Qy 550 TTGAGAGAGGTCAATAAACCAGGTGCAGGATATACACCAACAGTCCCGAACCCTTATAAC 609
```





CITY: Abbott Park  
 STATE: Illinois  
 COUNTRY: USA  
 ZIP: 60064  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
 MEDIUM TYPE: storage  
 COMPUTER: IBM  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: WordPerfect 5.1  
 CURRENT APPLICATION DATA: US/08/259,451  
 APPLICATION NUMBER: 436  
 FILING DATE: 20-JUN-1994  
 CLASSIFICATION: 436  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/086,415  
 FILING DATE: 01-JUL-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Daniel W. Collins  
 REGISTRATION NUMBER: 31,912  
 REFERENCE/DOCKET NUMBER: 5381.US.P1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (708) 937-6365  
 TELEFAX: (708) 938-2623  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 917 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 US-08-259-451-11

## Alignment Scores:

Pred. No.: 2,51e-51 Length: 917  
 Score: 629.50 Matches: 274  
 Percent Similarity: 36.6% Conservative: 115  
 Best Local Similarity: 25.8% Mismatches: 425  
 Query Match: 10.2% Indels: 249  
 DB: 2 Gaps: 38

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-08-259-451-11 (1-917)

QY 346 CCACAGTTATTCAGTGAAGCGCAGTGCACACCACTGTCAGTCAAGAGTACCCCTTG 405  
 Db |||||  
 QY 38 ProArgThrProSerProThrAlaProGlyGlyProIleSerPheLys----- 53  
 QY 406 AGTAAAGAGCTCAAGAGGAAATTCGGCGCATGTCCAAAGATTATCCACAGGGCATC 455  
 Db |||||  
 QY 54 ---ProGluArgLeuGlnAlaLeuAsnAspLeuValSerLysAlaLeuGluAlaGlyHis 72  
 QY 466 CTAGTTCCTGTCGAATCTCCTGGAATCTCCCTGCTACCGGTTAGAAAGCCTGGGACT 525  
 Db |||||  
 QY 73 IleGluProTyrSerGlyProGlyAsnAsnProValPheProValLysLysPro---Asn 91  
 QY 526 AATGACTATGACACGATACAGGACTTCAGAGAGGTCAATAAAGCGGTGCAGGATATACAC 585  
 Db |||||  
 QY 92 GlyLysTrpArgPheIleHisAspLeuArgAlaThrAsnAlaIleAlaThrThrLeuThr 111  
 QY 586 CCAACAGTCCGAAACCTTATACCTCTGTGTGCTCTCCACCCCAACCGAGCTGGTAT 645  
 Db |||||  
 QY 112 SerProSerProGlyProProAsp---LeuThrSerLeuProThrAlaLeuProHisLeu 130  
 QY 646 ACAGTATTGAGCTTAAAGGATGCTCTCTGCTGAGATTACACCCCACTAGCAACCA 705  
 Db |||||  
 QY 131 GlnThrIleAspLeuThrAspAlaPhePheGlnIleProLeuProLysGlnPheGlnPro 150  
 QY 706 CTTTTCCTTCGAATGAGAGATCCA---GGTACGGGAAGAACCGGGCAGCTCACCTGG 762  
 Db |||||  
 QY 151 TyrPheAlaPheThrIleProGlnProCysAsnTyrGlyProGlyThrArgTyrAlaTip 170  
 QY 763 ACCGAGTCCCAAGGTTCAAGAACTCCCGACCATCTTTACGAAAGCCCTTACACAGA 822  
 Db |||||  
 QY 171 ThrValLeuProGlnGlyPheLysAsnSerProThrLeuPheGluGlnLeuAlaAla 190

QY 823 GACCTGGCCAACTTCAGGATCCAAACCCCTCAGTGACCTCTCCAGTACGTGATGAC 882  
 Db |||||  
 QY 191 ValLeuAsnProMetArgLysMetPheProThrSerThrIleValGlnTyrMetAspAsp 210  
 QY 883 CTGCTTCTGGGGAGGCCACAAACAGGACTGCTTAGAAGCACGAGACGACTACTGCTG 942  
 Db |||||  
 QY 211 IleLeuLeuAlaSerProThrAsnLysGlnGlnLeuSerGlnLeuThrLeuGln 230  
 QY 943 GAATTGCTGACCTAGGCTACAGAGCTCTGCTAAGAAGGCCCAAGATTTCAGGAGAGAG 1002  
 Db |||||  
 QY 231 AlaLeuThrThrHisGlyLeuProIleSerGlnGlnLysThrGlnArgThrProGlyGln 250  
 QY 1003 GTAACATACTTGGGTACAGTTCCTGGGAGCGGCGAGCTGCTGACGAGGACCGAAG 1062  
 Db |||||  
 QY 251 IleArgPheLeuGlyGlnValIleSerProAsnHis-----IleThrTyrGluSer 267  
 QY 1063 AAAACTGTAGTCAGATACCGGCCCAACACAGCCAAACAAATAGAGAGATTTTGGGG 1122  
 Db |||||  
 QY 268 ThrProAlaIleProIleLysSerGlnTrpThrLeuThrGluLeuGlnValIleLeuGly 287  
 QY 1123 ACAGCTGGATTTCAGACTGTGATCCCG-----GGTTTGGCACTTACCA 1170  
 Db |||||  
 QY 288 GluIleGlnTrpValSerLysGlyThrProIleLeuArgLysHisLeuGlnSerLeuTyr 307  
 QY 1171 GCCCCACTCTACCCGCTAACCAAGAAAGGGAATTCTCTGGGCTCTCTGAGCACCA 1230  
 Db |||||  
 QY 308 SerAlaLeuHisGlyTyrArgAspProArgAlaCysIleThrLeuThrProGlnGlnLeu 327  
 QY 1231 AAGCATTTGATGCTATCAAAAGGCCCTCTGAGCGCACCTGCTGGGCCCTCCCTGAC 1290  
 Db |||||  
 QY 328 HisAlaLeuHisAlaIleGlnGlnAlaLeuGlnHisAsnCysArgGlyArgLeu---Asp 346  
 QY 1291 GTAACATAACCCCTTACCTTTATGTCGATGAGCGTAAGGAGTAGCCCGGAGGTTTA 1350  
 Db |||||  
 QY 347 ProThrLeuProLeuLeuGlyLeuIleSerLeuSerThrSerGlyThrThrSerValIle 366  
 QY 1351 ACCCAACCCCTAGGACCATGGAGAGACCTGTGCTGCTTACCTGTCAAGAGAGCTCGATCCT 1410  
 Db |||||  
 QY 367 PheGlnProLysGlnAsnTrp-----ProLeuAlaTrpLeuHisThrProHisProPro 384  
 QY 1411 GTAGCC---AGTGGTGGCCCATATGCTGAAGGCTATCCGAGTGTGGCCATCTGCTG 1467  
 Db |||||  
 QY 385 ThrSerLeuCysProTrpGlyHisLeuLeu-----AlaCysThrIle 398  
 QY 1468 AAGGACGCTGACAAATTTGACTTTGGGACAGAAATAACTGTAATA----- 1512  
 Db |||||  
 QY 399 LeuThrLeuAspLysTyrThrLeu---GlnHisTyrGlyLeuLeuCysGlnSerPheHis 417  
 QY 1513 -----GCCCCCAT----- 1521  
 Db |||||  
 QY 418 HisAsnMetSerLysGlnAlaLeuCysAspPheLeuArgAsnSerProHisProSerVal 437  
 QY 1522 -----GCATTGGAGAAATCGTTCCGGACGCCGCCACCGCA 1557  
 QY 438 GlyIleLeuIleHisMetGlyArgPheHisAsnLeuGlySerGlnProSerGlyPro 457  
 QY 1558 TGGATGACCAACGCCCGCATGACCCACTATCAAGGCTGCTTCTCAGAGAGGGTCAAG 1617  
 Db |||||  
 QY 458 TrpLysThr-----LeuLeuHisLeuProThrLeuLeu----- 469  
 QY 1618 TTGCTCTCCACCGCTCTCAACCTGCGCCTCTTCTGCTGAGAGACTGATGAACCA 1677  
 Db |||||  
 QY 469 -----GlnGluProArgLeuLeuArgProIlePheThrLeuSerPro 482  
 QY 1678 GTGACTCATGATTCATCAACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACA 1737  
 Db |||||  
 QY 483 ValValLeuAspThrAlaProCysLeu----- 491  
 QY 1738 GACATACCGCTGACTGGAGAGTGTAACTGTTTCACTGACCGAAGCAGCTATGTGGTG 1797  
 Db |||||  
 QY 492 -----PheSerAspGlySerPro----- 497

```
QY 1798 GAAGTAAGAGGTGGCTGGCGCGGTGGTGGACGGGACCCGACGACTCTGG----- 1851
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1852 ---GCCAGCACCTGCGC-----GAAGAACTTCAGCACAAAGAGGTGAGCTCATGGCC 1902
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 513 AspileThrProLeuProHisGluThrAsnSerAlaGlnLysGlyGluLeuLeuAla 532
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1903 CTCAGCGAAGCTTTCGGCTGGCGGAA---GGGAAATCCATAAACATTATTACGGACAGC 1959
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 533 LeuIleTyGlyLeuArgAlaAlaLysProTrpProSerLeuAsnIlePheLeuAspSer 552
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1960 AGGTATGCTTTCGACTGCACAC-----GTACATGGGGCGCATCTATAAACAAAGGGGG 2013
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 553 LysTyLeuIleLysTyLeuHisSerLeuAlaIleGlyAlaPhe----- 567
QY 2014 TTGCTTACCTCAGCAGGAGGAATAAGAACAAAGAGGAATAATTCTAAGCCTATTAGAA 2073
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 568 LeuGlyThrSerAlaHisGlnThrLeuGln-----AlaAlaLeuPro 581
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2074 GCGGTACATTTACCAAAAGCTAGCTATTATATACACTGTCTCTGGACATCAAGAGCTTAA 2133
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 582 ProLeuLeuGlnGlyLysThrIleTyLeuHisHisValArgSerHisThrAsnLeuPro 601
QY 2134 GATCTCATATCCAGAGGAAACACAGATGCTGCACGGGTTCACAGCAGGCCAGCGGT 2193
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 602 AspProIleSerThrPheAsnGluTyThrAspSerLeuIle----- 615
QY 2194 GTTAACCTTCTGCTTATATAAGAAATGCCAAAGCCCCAGAACCCAGACGACGTACACC 2253
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 616 -----ValAlaProLeuValProLeu----- 622
QY 2254 CTAGAAGACTGGCAAGAGATAAAAAGATAGACCGTTCTCTGAGACTCCCGAAGGGACC 2313
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 623 -----ThrProGlnGly----- 626
QY 2314 TGCTATACCTCAGATGGGAAGAAATCTCTCCAGAAATCTGCCACAAAGAGGTTAGATATGTCCA 2373
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 626 ----- 626
QY 2374 CAGATACATCGTCTAACCACCTAGGAACCTAAACACCTGCAGAGTTGGTCAGAACATCC 2433
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 627 ---LeuHisGlyLeuThrHisCysAsnGlnArgAlaLeuValSerPheGlyAlaThr--- 644
QY 2434 CCTTATCATGTTCTGAGGCTACCAGGAGTGCCTGACTCGGTGGTCAACATTTGTGCCC 2493
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 645 -----ProlysGluAlaLysSerLeuValGlnThrCysHisThr 657
QY 2494 TGCCAGCTGGTTAATGCTAATCTCTCCAGAATCGCTCCAGGAGAGAGACTAAGGGGAAGC 2553
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 658 CysGlnIleIleAsnSerGln---HisHisMetProGlnGlnHisIleArgArgGlyLeu 676
QY 2554 CACCCAGCGCTCACTGGGAAGTGGACTTCATGAGGTAAAGCCCGGCTAAATACGGAAAC 2613
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 677 LeuProAsnHisIleIleTrpGlnGlyAspValThrHisTyLysTyLysLysTyLys 694
QY 2614 AAATACCTATTG---GTTTTGTAGACACCTTTTCAGGATGGGTAGAGGCTTATCCTACT 2670
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 695 LysTyCysLeuHisValTrpValAspThrPheSerGlyAlaValSerValSerCysLys 714
QY 2671 AAGAAAGAGACTTCAACCGTGTGCTAAAAAATACTGGAAGAAATTTTCCAGATT 2730
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 715 LysLysGluThrSerCysGluThrIleSerAlaPheLeuGlnAlaIle---SerLeuLeu 733
QY 2731 GGAATACCTAAGGTAAAGGTACAGCAATGGTCCAGCTTTTGTGGCCAGGTAAAGTCAG 2790
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 734 GlyLysProLeuHisIleAsnThrAspAsnGlyProAlaPheLeuSerGlnGluPheGln 753
QY 2791 GGACTGGCCAAGATATTCGGGATTGATGGAACTGGCAATGTGCATACAGACCCCAAGC 2850
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 754 GluPheCysThrSerTyHisIleLysHisSerThrHisIleProTyAsnProThrSer 773
QY 2851 TCAGGACAGGTAGAGAGGATGAATAGAACCATTTAAAGAGACCTTACTTAAAA----- 2901
```

```
Db 774 SerGlyLeuValGluArgThrAsnGlyIleIleLysAsnLeuLeuAsnLysTyLeuLeu 793
QY 2902 -----TTGACCGGGAGAGACTGGCGTTAATGATGGATAGCTCTCTCCCTTT 2949
Db 794 AspCysProAsnLeuProLeuAspAsnAlaIleAsnLys-----AlaLeuTrpThrLeu 811
QY 2950 GTGCTTTTGGGTAGGAACACCCCTGGACAGTTTGGGCTGACCCCTATGAATTACTC 3009
Db 812 AsnGlnLeuAsnValMetAsnProSerGlyLys-----ThrArgTrpGlnIleHis 828
QY 3010 TACGGGGGACCCCCCAATTGGTAGAAATTCTTCTGTACATAGTGTGACGTCTGCTT 3069
Db 829 HisSer-----ProProLeuProProIleProGluAlaSerThrProProLysProPro 846
QY 3070 TCCAGAGCTTTGTTCTCTAGGCTCAAGGCATTGAGTGGGTGAGACACAGCGTGGAGG 3129
Db 847 SerLysTrpPheTyTyTyLysLeuProGlyLeu----- 857
QY 3130 CAACCTCCGGGAGGCTACTCAGGAGGAGAGACTTGCAGATCCCACTCGTTCCAAAGTG 3189
Db 857 ----- 857
QY 3190 GGAGATTACGTCTACGTGTAGACGCCACCCTGCAGGAAACCTCGAGACTCGGTGGAAGGCG 3249
Db 858 -----ThrAsnGlnArgTrpLysGly 864
QY 3250 CCTTATCTCGTACTTTTGACACACCAACCGCT-----GTGAAAGTCGAAGGAATCTCC 3303
Db 865 ProLeuGlnSerLeuGlnGluAlaAlaGlyAlaAlaLeuLeuSerIleAspGlyPhePro 884
QY 3304 ACCTGGATCCATGCATCCACGTTAAACCGCG-----CCACCTCCCGATTTCGGGGTGG 3357
Db 885 ArgTrpIleProTrpArgPheLeuLysLysAlaAlaCysProArgProAspAlaSerGlu 904
QY 3358 AAAGCCGAA 3366
Db 905 ProAlaGlu 907
```

Search completed: February 14, 2006, 16:17:52  
Job time : 226.767 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 16:07:51 ; Search time 86.3713 Seconds  
(without alignments)  
3323.427 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_2307\_5741

Perfect score: 6183

Sequence: 1 ATGGGTGCCAGGCGCAACA.....CTGTCAATACCTCTCAGAC 3435

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSSWEB.spool/US10723552/runat\_14022006\_125148\_13205/app.query.fasta\_1  
-DB=Published Applications AA Main -QWTF=fastan -SUFFIX=rapbm -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blisum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs02p  
-USER=US10723552 @CGN 1.1 602 @runat\_14022006\_125148\_13205 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -BSPLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA Main:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3808.5	61.6	1199	4	US-10-677-558-2
2	2491	40.3	678	5	US-10-827-498-2
3	2487.5	40.2	716	3	US-09-845-157-2
4	2487.5	40.2	716	4	US-10-661-819-2
5	2007	32.5	471	4	US-10-448-871A-31
6	1867	30.2	471	4	US-10-448-871A-29
7	1862	30.1	471	4	US-10-448-871A-27
8	1850.5	29.9	470	4	US-10-448-871A-28
9	1846	29.9	471	4	US-10-448-871A-30
10	1794	29.0	1438	5	US-10-450-763-42377
11	1784	28.9	473	4	US-10-448-871A-25

12	1768.5	28.6	472	4	US-10-448-871A-26
13	1431.5	23.2	1267	5	US-10-450-763-33781
14	1415.5	22.9	768	2	US-08-979-847-89
15	1415.5	22.9	768	4	US-10-114-104-89
16	1408.5	22.8	817	5	US-10-450-763-40498
17	1406.5	22.7	896	5	US-10-450-763-34798
18	1390.5	22.5	683	2	US-08-979-847-198
19	1390.5	22.5	683	2	US-08-979-847-200
20	1390.5	22.5	683	4	US-10-114-104-198
21	1390.5	22.5	683	4	US-10-114-104-200
22	1387.5	22.4	683	2	US-08-979-847-208
23	1387.5	22.4	683	2	US-08-979-847-210
24	1387.5	22.4	683	4	US-10-114-104-208
25	1387.5	22.4	683	4	US-10-114-104-210
26	1363.5	22.1	821	5	US-10-450-763-52800
27	1345.5	21.8	872	5	US-10-450-763-42647
28	1345.5	21.8	1288	5	US-10-450-763-34793
29	1345.5	21.8	1288	5	US-10-450-763-42636
30	1331.5	21.5	654	2	US-08-979-847-91
31	1331.5	21.5	654	4	US-10-114-104-91
32	1330.5	21.5	1067	5	US-10-450-763-38408
33	1280	20.7	805	5	US-10-450-763-47721
34	1268.5	20.5	997	5	US-10-450-763-39301
35	1248.5	20.2	1253	5	US-10-450-763-57733
36	1208	19.5	694	5	US-10-450-763-37899
37	1181	19.1	1577	5	US-10-450-763-36834
38	1181	19.1	1577	5	US-10-450-763-57482
39	1101.5	17.8	930	5	US-10-450-763-40953
40	1101.5	17.8	1142	5	US-10-450-763-35673
41	1093.5	17.7	1001	5	US-10-450-763-47711
42	1047	16.9	614	5	US-10-450-763-39777
43	1047	16.9	614	5	US-10-450-763-50364
44	1042	16.9	552	5	US-10-450-763-39780
45	1042	16.9	552	5	US-10-450-763-42631

#### ALIGNMENTS

##### RESULT 1

US-10-677-558-2  
; Sequence 2, Application US/10677558  
; Publication No. US20040096972A1  
; GENERAL INFORMATION:  
; APPLICANT: AUDIT, Muziel  
; APPLICANT: COSSET, Francois-Loic  
; TITLE OF INVENTION: CHIMERIC PLASMID COMPRISING A REPLICATIVE RETROVIRAL GENOME AND  
; FILE REFERENCE: 1759.135  
; CURRENT APPLICATION NUMBER: US/10/677,558  
; CURRENT FILING DATE: 2003-10-02  
; PRIOR APPLICATION NUMBER: PCT/FR02/03934  
; PRIOR FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: FR 0114976  
; PRIOR FILING DATE: 2001-11-20  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 2  
; LENGTH: 1199  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: product of pol gene of pAM plasmid  
US-10-677-558-2

Alignment Scores: 6.12e-300 Length: 1199  
Score: 3808.50 Matches: 722  
Best Local Similarity: 76.2% Conservatives: 160  
Best Local Similarity: 62.3% Mismatches: 245  
Query Match: 61.6% Indels: 31  
DB: 4 Gaps: 11

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-10-677-558-2 (1-1199)



QY	3253	TATCTCGTACTTTTGACACACCAACGGCTGTGAAATCGAAGAAATCTCCACCTGGATC	3312
Db	1142	TyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleAlaAlaTrpIle	1161
QY	3313	CATGCATCCACAGTTAAACCGCGCCACCT-----CCCGATTCCGGG-----TGG	3357
Db	1162	HisAlaAlaHisValIysAlaAlaAspThrGluSerGlyProSerSerGlyArgThrTrp	1181
QY	3358	AAAGCCGAAAGACTGAAATCCCTTAAAGCTTCGCCTCCATCGCGTGGTTCCCT	3411
Db	1182	ArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThrArgGlySerPro	1199

RESULT 2

US-10-827-498-2

; Sequence 2, Application US/10827498

; Publication No. US20050232934A1

; GENERAL INFORMATION:

; APPLICANT: Ambion, Inc.

; APPLICANT: Chen, Lianling

; APPLICANT: Satterquist, Robert

; APPLICANT: Latham, Gary

; TITLE OF INVENTION: Recombinant Reverse Transcriptases

; FILE REFERENCE: AMBI:1001

; CURRENT APPLICATION NUMBER: US/10/827,498

; CURRENT FILING DATE: 2004-04-19

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 2

; LENGTH: 678

; TYPE: PRT

; ORGANISM: Moloney murine sarcoma virus

US-10-827-498-2

Alignment Scores:

Pred. No.:	5.37e-193	Length:	678
Score:	2491.00	Matches:	462
Percent Similarity:	82.6%	Conservative:	95
Best Local Similarity:	68.5%	Mismatches:	107
Query Match:	40.3%	Indels:	10
DB:	5	Gaps:	4

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-10-827-498-2 (1-678)

QY	208	TTGACCTCAATTAGATGACGATATCGACTATACTCTCCCTAGTAAGCTGATCAA	267
Db	1	MetThrLeuAsnIleGluAspGluTyrArgLeuHisGluThrSerLysGluProAspVal	20
QY	268	AATATA---CAATTCTGGTTGGAAACAGTTTCCCAAGCCTGGGCAGAAACCGCAGGGATG	324
Db	21	SerLeuGlySerThrTrpLeuSerAspPheProGlnAlaTrpIleGluThrGlyGlyMet	40
QY	325	GGTTTGGCAAGCAAGTTTCCCCCAACAGTATTTCAACTGAAGCCAGTCACACCAAGTG	384
Db	41	GlyLeuAlaValArgGlnAlaProLeuIleIleProLeuLysAlaThrSerThrProVal	60
QY	385	TCAGTCAGACAGTACCCCTGTGATTAAGAGCTCAAGAGGAATTCCGCCGATGTCCAA	444
Db	61	SerIleLysGlnTyrProMetSerGlnGluAlaArgLeuGlyIleLysProHisIleGln	80
QY	445	AGATAATCAACACAGGGCATCTCTAGTTCCTCCAATCTCCCTGGAAATCTCCCTCGCTA	504
Db	81	ArgLeuLeuAspGlnGlyIleLeuValProCysGlnSerProTrpIleThrProLeuLeu	100
QY	505	CCGGTTAGAAGCCTGGGACTAATGACTATCGACCAAGTACAGGACTTGAGAGAGGTCAAT	564
Db	101	ProValLysLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsn	120
QY	565	AAACGGGTGCAGATATACACCAACAGTCCCAACCCCTATATACCTCTTGTTGTGCTCTC	624
Db	121	LysArgValGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerGlyLeu	140
QY	625	CCACCCCAACGGGAGCTGGTATACAGTATTGGACTTAAAGGATCGCTTCTTCTCGCTGAGA	684









```
QY 2122 CAGAAAGCTAAGATCTCATATCCAGAGAAACCAAGATGGCTGACCGGTTGCCAAGCAG 2181
|||||
Db 672 GlnlyGlyHisSerAlaGluAlaArgGlyAsnArgMetAlaAspGlnAlaAlaArglys 691
|||||
QY 2182 GCAGCC 2187
|||||
Db 692 AlaAla 693

RESULT 5
US-10-448-871A-31
; Sequence 31, Application US/10448871A
; Publication No. US20040093169A1
; GENERAL INFORMATION:
; APPLICANT: Mizushima, Yoshiyuki
; APPLICANT: Sakaguchi, Kengo
; APPLICANT: Sugawara, Fumio
; TITLE OF INVENTION: A METHOD FOR DESIGNING A MOLECULAR STRUCTURE OF AN INHIBITOR FOR
; FILE OF INVENTION: AN ENZYME
; FILE REFERENCE: 03316/LH
; CURRENT APPLICATION NUMBER: US/10/448,871A
; CURRENT FILING DATE: 2003-05-29
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 31
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Gibbon endogenous virus
US-10-448-871A-31

Alignment Scores:
Pred. No.: 9,73e-154 Length: 471
Score: 2007.00 Matches: 367
Percent Similarity: 88.7% Conservative: 50
Best Local Similarity: 78.1% Mismatches: 53
Query Match: 32.5% Indels: 0
DB: 4 Gaps: 0

US-10-723-552-3_copy_2307_5741 (1-3435) x US-10-448-871A-31 (1-471)

QY 325 GGTGTTGCAAGCAAGTATCCCAACAGTTATTTCAACTGAGGCCAGTGCACACCAAGT 384
|||||
Db 1 GlyLeuAlaAsnGlnValProValValGluLeuArgSerGlyAlaSerProVal 20
|||||
QY 385 TCAGTCAGACAGTACCCCTTGAGTAAGAAGCTCAAGAAGAAATTCGGCCGCGCATGTCCAA 444
|||||
Db 21 AlaValargGlnTyProMetSerlyGluAlaargGluGlyIleArgProHisIleGln 40
|||||
QY 445 AGATTATCCACAGGGGATCCTAGTTCCTGTCCAAATCTCCCTGGAATACCTCCCTGCTA 504
|||||
Db 41 LysPheLeuAspLeuGlyValLeuValProCysArgSerProTyrAsnThrProLeuLeu 60
|||||
QY 505 CCGGTTAGAAAGCCTGGGACTTAATGACTATCGACAGTACAGGACTTGAGAGAGTCAAT 564
|||||
Db 61 ProVallyLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluIleAsn 80
|||||
QY 565 AAACGGGTGAGGATATACACCCACAGTCCCGAACCTTATACCTCTTGTGTGCTCTC 624
|||||
Db 81 LysArgValGlnAspIleHisProThrValProAsnProTyrAsnLeuLeuSerSerLeu 100
|||||
QY 625 CCACCCCAACGGAGCTGGTATACAGTATGTGACTTAAAGGATGCTTTCTTCTGCTGAGA 684
|||||
Db 101 ProProSerTyrThrTyrSerValLeuAspLeuLysAspAlaPhePheCysLeuArg 120
|||||
QY 685 TTACACCCCACTAGCCCAACCACTTTTGGCTTCGAATGGAGATCCAGATCCGGGAAGA 744
|||||
Db 121 LeuHisProAsnSerGlnProLeuPheAlaPheGluThrLysAspProGluLysGlyAsn 140
|||||
QY 745 ACCGGGAGCTCAGCTGACCCGACTCCCAAGGTTCAAGAACTCCCGACCATCTTT 804
|||||
Db 141 ThrGlyGlnLeuThrTyrThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPhe 160
|||||
QY 805 GACGAAGCCCTACACAGAGACCTGGCCAACCTTCAGGATCCACACCCCTCAGGTGACCCCTC 864
|||||
```

```
Db 161 AspGluAlaLeuHisArgAspLeuAlaProPheArgAlaLeuAsnProGlnValValLeu 180
|||||
QY 865 CTCAGATACGTGGATGACCTCTCTCGCGGAGCCCAACAGGACTGCTTGAAGGC 924
|||||
Db 181 LeuGlnTyValAspAspLeuLeuValAlaAlaProThrTyrGluAspCysLysGly 200
|||||
QY 925 ACGAAGCAGCTACTGCTGGAATTCCTGACCTAGGCTACAGAGCTCTGCTAAGAAGGCC 984
|||||
Db 201 ThrGlnLysLeuLeuGlnGluLeuSerLysLeuGlyTyrArgValSerAlaLysLysAla 220
|||||
QY 985 CAGATTTGAGGAGAGAGGTTAAATATCTTGGGGTACAGTTTCCGGGAGCGGACGATGG 1044
|||||
Db 221 GlnLeuCysGlnArgGluValThrTyrLeuGlyTyrLeuLysGluGlyLysArgTrp 240
|||||
QY 1045 CTGACGAGGACGCGAAGAACTGTAGTCCAGATACCGGCCCAACACACAGCCCAACAA 1104
|||||
Db 241 LeuThrProAlaArgLysAlaThrValMetLysIleProValProThrProArgGln 260
|||||
QY 1105 ATGAGAGAGTTTTTGGGACAGCTGGATTTTGACAGCTGTGGATCCCGGGTTTCCGACC 1164
|||||
Db 261 ValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaSer 280
|||||
QY 1165 TTAGACGCCCTCTCTACCCGCTTAACCAAGAAAGGGGAATTCCTCTGGGCTCCTGAG 1224
|||||
Db 281 LeuAlaAlaProLeuTyrProLeuThrLysGluSerIleProPheIleTrpThrGluGlu 300
|||||
QY 1225 CACGAGAAGCATTGTATGCTATCAAAAGCCCTGTGAGCGCACCTGCTCTGCGCCCTC 1284
|||||
Db 301 HisGlnGlnAlaPheAspHisIleLysIleAlaLeuLeuSerAlaProAlaLeuAlaLeu 320
|||||
QY 1285 CCTGACGTAATAACCCCTTTTACCTTTATGTGTGATGAGGTAGGAGTAGCCCGGGA 1344
|||||
Db 321 ProAspLeuThrLysProPheThrLeuTyrIleAspGluArgAlaGlyValAlaArgGly 340
|||||
QY 1345 GTTTTAACCAACCCCTAGGACCATGAGAAAGACCTCTCGCTTACCTGTCAAAGAGCTC 1404
|||||
Db 341 ValLeuThrGlnThrLeuGlyProTyrArgProValAlaTyrLeuSerLysLysLeu 360
|||||
QY 1405 GATCCTGTAGCAGTGTGCGCCATATGCTGAAGCTATCGCAGCTGTGGCCATCTG 1464
|||||
Db 361 AspProValAlaSerGlyTyrProThrCysLeuLysAlaValAlaAlaValAlaLeuLeu 380
|||||
QY 1465 GTCAAGACGCTGACAAATTCAGCTTTGGGACAGATAACTGTAATAGCCCCCATGCA 1524
|||||
Db 381 LeuLysAspAlaAspLysLeuThrLeuGlyGlnAsnValThrValIleAlaSerHisSer 400
|||||
QY 1525 TTGAGAACATCGTTTGGCAGCCCCCAGACCGATGGATGACCAACGCCCGCATGCCAC 1584
|||||
Db 401 LeuGluSerIleValArgGlnProProAspArgTrpMetThrAsnAlaArgMetThrHis 420
|||||
QY 1585 TATCAAGCCTGCTTCTCACAGAGAGGTCACGTCGCTCCACGCGGCTCTCAACCT 1644
|||||
Db 421 TyrGlnSerLeuLeuLeuAsnGluArgValSerPheAlaProProAlaValLeuAsnPro 440
|||||
QY 1645 GCACACTCTCTGCTGAAGAGACTGATGAACCAAGTACTCATGATTCGCTCAACTATTG 1704
|||||
Db 441 AlaThrLeuLeuProValGluSerGluAlaThrProValHisArgCysSerGluIleLeu 460
|||||
QY 1705 ATTGAGGAGACTGGGTCCGCAAGGACCTT 1734
|||||
Db 461 AlaGluGluThrGlyThrArgArgAspLeu 470
|||||
```

## RESULT 6

```
US-10-448-871A-29
; Sequence 29, Application US/10448871A
; Publication No. US20040093169A1
; GENERAL INFORMATION:
; APPLICANT: Mizushima, Yoshiyuki
; APPLICANT: Sakaguchi, Kengo
; APPLICANT: Sugawara, Fumio
; TITLE OF INVENTION: A METHOD FOR DESIGNING A MOLECULAR STRUCTURE OF AN INHIBITOR FOR
; TITLE OF INVENTION: AN ENZYME
```

```
; FILE REFERENCE: 03316/LH
; CURRENT APPLICATION NUMBER: US/10/448,871A
; CURRENT FILING DATE: 2003-05-29
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 29
; LENGTH: 471
; TYPE: PRT
; ORGANISM: AKV murine leukemia virus
US-10-448-871A-29

Alignment Scores:
Pred. No.:      2,31e-142      Length:      471
Score:          1867.00      Matches:     339
Percent Similarity: 85.6%      Conservative: 65
Best Local Similarity: 71.8%      Mismatches: 66
Query Match:    30.2%      Indels:      2
DB:              4      Gaps:         2

US-10-723-552-3_COPY_2307_5741 (1-3435) x US-10-448-871A-29 (1-471)

QY 325 GGTGGCAAGCAAGTTCCTCCCAAGTATTCACTGAAGGCGCAGTGCACACCACTG 384
Db 1 GlyLeuAlaValArgGlnAlaProLeuLysAlaThrSerThrProVal 20
QY 385 TCAGTCACAGCAGTACCCCTTGAGTAAAGAGCTCAAGAGGAAATTCGGCGCATGTCCAA 444
Db 21 SerIleLysGlnTyrProMetSerGlnGluAlaLysLeuGlyIleLysProHisIleGln 40
QY 445 AGATTAAATCCAAACAGGGCATCTAGTCTCTGTGCCAATCTCCCTGGAACTCCCTGTCTA 504
Db 41 ArgLeuLeuAspGlnGlyIleLeuValProCysGlnSerProTyrAsnThrProLeuLeu 60
QY 505 CGGTTAGAAAGCCCTGGGACTAATGACTATCGACAGTACGAGCTTGAGAGAGTCAAT 564
Db 61 ProValLysLysProGlyThrAsnAspTyrArgProValGlnAspLeuArgGluValAsn 80
QY 565 AAACGGGTGCAGGATATACCCCAACAGTCCCGACCCCTTATACCTCTTGTCCTC 624
Db 81 LysArgValGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerGlyLeu 100
QY 625 CCACCCCAACGGAGCTGTGTATACAGTATTGGAATTAAGGATGCTCTTCTGCTGAGA 684
Db 101 ProProSerHisArgTyrThrValLeuAspLeuLysAspAlaPheCysLeuArg 120
QY 744 TTACACCCCACTAGCCCAACCACTTTTGGCTTCGAATGGAGAGATCCAGGTACGGGAAGA 744
Db 121 LeuHisProThrSerGlnProLeuPheAlaPheGluTyrArgAspProGlyMetGlyIle 140
QY 804 ACCGGGACCTCACAGACCTGCGCAACTTCAGATCCAAACCCCTCAGGTGACCCCTC 864
Db 161 AspGluAlaLeuHisArgAspLeuAlaAspPheArgIleGlnHisProAspLeuLeuLeu 180
QY 865 CTCACGTACGTGAGTACCTGCTTGGCGGAGCCCAACACAGGACTGCTTAGAAGGC 924
Db 181 LeuGlnTyrValAspAspIleLeuSerAlaAlaThrSerGluLeuAspCysGlnGlnGly 200
QY 925 ACGAGGCACTACTGCTGGAATGCTGACCTAGGCTACAGGCTCTGCTTAAGAGGCC 984
Db 201 ThrArgAlaLeuLeuLeuThrLeuGlyAsnLeuGlyTyrArgAlaSerAlaLysLysAla 220
QY 985 CAGATTTGCAGGAGAGAGTAACTATCTTGGGTACAGTTTGGCGGAGCGGCAGCATGG 1044
Db 221 GlnLeuCysGlnLysGlnValLysTyrLeuGlyTyrLeuLeuLysGluGlyGlnArgTyr 240
QY 1045 CTGACGGAGGCGCAGGAAGAACTATGTATCCAGATACCGGCGCCCAACACCAAGCAAA 1104
Db 241 LeuThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGln 260
```

```
QY 1105 ATGAGAGAGATTTTGGGACAGCTGGATTTTGCAGACTGTGTGATCCCGGGTGTCCGACC 1164
Db 261 LeuArgGluPheLeuGlyThrAlaGlyPheCysargLeuTyrIleProGlyPheAlaGlu 280
QY 1165 TTAGACGCCCACTCTATCCCGCTAACCAAGAAAGGGGAATTCCTCTGGGCTCCTGAG 1224
Db 281 MetAlaAlaProLeuTyrProLeuThrLysThrGlyThrLeuPheAsnTyrGlyProAsp 300
QY 1225 CACCAGAGGAGATTGATGCTATCAAAAGCCCTGTGAGCGCAGCTGCTCTGCGCCCTC 1284
Db 301 GlnGlnLysAlaTyrGlnGluIleLysGlnAlaLeuLeuThrAlaProAlaLeuGlyLeu 320
QY 1285 CCTGACGTAACTAAACCTTTTACCCTTTATGTGATGAGCGCTAAGGGAGTAGTCCGGGGA 1344
Db 321 ProAspLeuThrLysProPheGluLeuPheValAspGluLysGlnGlyTyrAlaLysGly 340
QY 1345 GTTTTAAACCCAAACCTTAGGACCATGAGAGAAGCCTGTGCGCTACCTGTCAAAGAGCTC 1404
Db 341 ValLeuThrGlnLysLeuGlyProTyrArgArgProValAlaTyrLeuSerLysLysLeu 360
QY 1405 GATCCTGTAGCCAGTGTGGCCCATATGCTGCTGAAGGCTATCGCAGCTGTGCGCCATCTG 1464
Db 361 AspProValAlaAlaGlyTyrProProCysLeuArgMetValAlaAlaIleAlaValLeu 380
QY 1465 GTCAAGGACGCTGACAAATTGACTTTGGGACAGAAATATAACTGTAATAGCCCCCATGCA 1524
Db 381 ThrLysaspAlaGlyLysLeuThrMetGlyGlnProLeuValIleLeuAlaProHisAla 400
QY 1525 TTGGAGAACACTGTTCCGAGCCCAACCGATGGATGATGACCAACGCGCCGATGACCCAC 1584
Db 401 ValGluAlaLeuValLysGlnProProAspArgTyrLeuSerAsnAlaArgMetThrHis 420
QY 1585 TATCAAGCCCTGCTCTCTC---ACAGAGAGGCTCAGTTCGCTCCACCAGCCGCTCTCAAC 1641
Db 421 TyrGlnAlaMetLeuLeuAspThrAspArgValGlnPheGlyProValValAlaLeuAsn 440
QY 1642 CCTGCCACTCTCTGCTGAAGAGACTGATGAACAGTGAATCATGATGATGATGATGATGAT 1701
Db 441 ProIleThrLeuLeuPro---LeuProGluGluGlyAlaProHisAspCysLeuGluIle 459
QY 1702 TTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACA 1737
Db 460 LeuAlaGluThrHisGlyThrArgProAspLeuThr 471

RESULT 7
US-10-448-871A-27
; Sequence 27, Application US/10448871A
; Publication No. US20040093169A1
; GENERAL INFORMATION:
; APPLICANT: Mizushima, Yoshiyuki
; APPLICANT: Sakaguchi, Kengo
; APPLICANT: Sugawara, Fumio
; TITLE OF INVENTION: A METHOD FOR DESIGNING A MOLECULAR STRUCTURE OF AN INHIBITOR FOR
; FILE REFERENCE: 03316/LH
; CURRENT APPLICATION NUMBER: US/10/448,871A
; CURRENT FILING DATE: 2003-05-29
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 27
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Friend murine leukemia virus
US-10-448-871A-27

Alignment Scores:
Pred. No.:      5,89e-142      Length:      471
Score:          1862.00      Matches:     341
Percent Similarity: 85.2%      Conservative: 61
Best Local Similarity: 72.2%      Mismatches: 68
Query Match:    30.1%      Indels:      2
DB:              4      Gaps:         2
```

Qy	325	GGTTTGGCAAGCAAGTTCCCCCACACTATTTCACATGAAGGCCAGTGCCACACCAGTG	384
Dd		:::      :::     :::     :::     :::     :::     :::     :::	
Dd	1	GlyLeuAlaValArgGlnAlaProLeuIlelleProLeuLysAlaThrSerThrProVal	20
Qy	385	TCACTGACAGCAGTAGTCCCTCGAGTAAGAAGCTCAAAGAAGAAATTCGGCCGCATGTCCA	444
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	21	SerileLysGlnTyProMetSerGlnGlualaArgLeuGlyIleLysProHisleGln	40
Qy	445	AGATAATCCAAACAGGGCATCTCTAGTTCTGCTGCCTCAATCTCCCTGGAAATCTCCCTGCTA	504
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	41	ArgLeuLeuAspGlnGlyIleLeuValProCysGlnSerProTrpAsnThrProLeuLeu	60
Qy	505	CCGGTTAGAAAGCTGGGACTAAATGACTATACACCAGTCCCGAACCTCTATAACCTCTTGTTGCTCTC	564
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	61	ProVallylYsProGlyThrAsnAspTryrArgProValGlnAspLeuArgGluValAsn	80
Qy	565	AAACGGGTGACAGGTATATACACCCCAACAGTCCCGAACCTCTATAACCTCTTGTTGCTCTC	624
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	81	LysArgValGluAspIleHisProThrValProAsnProTyraAsnLeuLeuSerGlyLeu	100
Qy	625	CCACCCCAACGGAGCTGGTATACAGTATTGGACTTAAAGGATGCCCTTCTTCCTGCCTGAGA	684
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	101	ProProSerHisGlnTrpTyThrValLeuAspLeuLysAspAlaPhePheCysLeuArg	120
Qy	685	TTACACCCCACTAGCCACCACTTTTTGCTTCGAATGGAGATACAGGTACCGGAAGA	744
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	121	LeuHisProThrSerGlnSerLeuPheAlaPheGluTrpArgProGluMetGlyIle	140
Qy	745	ACCGGGCAGCTCACCTGGACCCGACTGCCCCAAGGGTTCAAGAACTCCCCGACCATCTTT	804
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	141	SerGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPhe	160
Qy	805	GACGAAGCCTACACAGAGACTGGCCCACTTCAGGATCCCAACCTTCAGTGACCTC	864
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	161	AspGluAlaLeuHisArgAspLeuAlaAspPheArgilleGlnHisProAspLeuLeu	180
Qy	865	CTCCAGTACGTGGATGACCTGCTTCCTGGCGGAGCCCAACACAGGACTGCTTAGAAGC	924
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	181	LeuGlnTyralAspAspLeuLeuLeuAlalaThrSerGluLeuAspCysGlnGlnGly	200
Qy	925	ACGAAGCCTACTGCTGGAAATTCCTGACCTAGGCTACAGAGCTCTGCTAAGAAGGCC	984
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	201	ThrArgAlaLeuLeuGlnThrLeuGlyAspleuGlyTyraGlaSerAlaLysylSala	220
Qy	985	CAGATTTCAGGAGAGAGGTAAACATCTTGGGGTACAGTTTGGCGGACGGCAGCGATGG	1044
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	221	GlnileCysGlnLysGlnValLysetyrLeuGlytyrLeuLeuLysGluGlyGlnArgTrp	240
Qy	1045	CTGACGAGGCACGGAAGAAACGTAGTCCAGATACCGGCCCAACACACAGCCACAACAA	1104
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	241	LeuThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGln	260
Qy	1105	ATGAGAGAGTTTTTGGGACAGCTGGATTTTTGCAGACTGTGGATCCCGGGTTTGGCCACC	1164
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	261	LeuArgGluPheLeuGlyThralaGlyPheCysargLeuTrpIleProGlyPheAlaGlu	280
Qy	1165	TTAGCAGCCCACTCTACCGCTAACCAAGAAAAGGGGAATTCCTCTGGGCTCTGTAG	1224
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	281	MetAlaalaProLeutyProLeuThrLysThrGlyThrLeuPheGluTrpGlyProAsp	300
Qy	1225	CACCAAGGCATTTGATGCTATCAAAAAGSCCTGTGTGACGCGACCTGCTCTGCCCTC	1284
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	301	GlnGlnLysAlatyGlnGluIleLysGlnAlaLeuLeuThrAlaProAlaLeuGlyLeu	320
Qy	1285	CCTGACCTAACTAACCCCTTTACCCCTTTATGTGGATGAGCTGAAGGAGTAGCCGGGGA	1344
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	321	ProAspLeuThrLysProPheGluLeuPheValaspGlnLysGlnGlyTyralaLysGly	340
Qy	1345	GTTTTAAACCAACCCCTAGGACCATGAGAGAGACTCTGCGCTTACCTGTCAAGAAGCTC	1404
Dd		:::     :::     :::     :::     :::     :::     :::     :::	
Dd	341	ValLeuThrGlnLysLeuGlyVProTrpArgArgProValalatyLeuSerLysLysLeu	360

[illegible]

Db 81 LysArgValGluAspIleHisProThrValProAsnProTyrAsnLeuSerGlyLeu 100  
Qy 625 CCACCCCAACGAGCTGTATACAGTATGAGCTTAAAGATGCTCTTCTGCTGAGA 684  
Db 101 ProProSerHisGlnTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArg 120  
Qy 695 TTACACCCACTAGCCCAACCACTTTTCCCTTCGATCGAGATCCAGGTACGGGAGA 744  
Db 121 LeuHisProThrSerGlnProLeuPheAlaPheGluTrpArgProGluMetGlyIle 140  
Qy 745 ACCGGCAGCTCACCTGACCGCTGCGCCCAAGGTTCAAGAACTCCCGCCACCATCTTT 804  
Db 141 SerGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPhe 160  
Qy 805 GACGAAGCCCTACACAGAGACTGCGCCAACTTCAGGATCCAAACACCTTCAGGTGACCCCTC 864  
Db 161 AspGluAlaLeuHisArgAspLeuAlaAspPheArgIleGlnHisProAspLeuIleLeu 180  
Qy 865 CTCAGTACGTGGATGACTGCTGCTGCGGGAGGCCAACACAGGACTGCTTAGAGGC 924  
Db 181 LeuGlnTyrValAspAspLeuLeuAlaAlaThrSerGluLeuAspCysGlnGlnGly 200  
Qy 925 ACGAAGGCACCTACTGCTGGAATTGCTCACCTAGCTACAGAGCTCTGCTAAAGAAGGCC 984  
Db 201 ThrArgAlaLeuLeuGlnThrLeuGlyAsnLeuGlyTyrArgAlaSerAlaLysAla 220  
Qy 985 CAGATTTTCAGGAGAGAGTAACTATCTGGGGTACAGTTTTCGGGACGGCAGCGATGG 1044  
Db 221 GlnIleCysGlnLysGlnValLysTyrLeuGlyTyrLeuLeuLysGluGlyGlnArgTrp 240  
Qy 1045 CTGACGGAGGCACGGAAGAACTGTAGTACAGATACCGGCCCCACACAGCCCAACAA 1104  
Db 241 LeuThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGln 260  
Qy 1105 ATGAGAGAGTTTTCGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTTCGGACC 1164  
Db 261 LeuArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGlu 280  
Qy 1165 TTACAGCCCCCACTACCCGCTAACCAAGAAAGGGGAATTCTCTCGGCTCCTGAG 1224  
Db 281 MetAlaAlaProLeuTyrProLeuThrLysThrGlyThrLeuPheAsnTrpGlyProAsp 300  
Qy 1225 CACGAGAAGGCATTTGATGCTATCAAAAGGCCCTGTGAGCGCACCTGCTCGCCCTC 1284  
Db 301 GlnGlnLysAlaTyrGlnGlnLysGlnAlaLeuLeuThrAlaProAlaLeuGlyLeu 320  
Qy 1285 CTGACGTAACTAAACCCCTTTACCCCTTTATGTGATGAGCGTAAAGGGAGTAGCCCGGGA 1344  
Db 321 ProAspLeuThrLysProPheGluLeuPheValAspGluLysGlnGlyTyrAlaLysGly 340  
Qy 1345 GTTTTAAACCCCAACCCCTAGGACCATGAGAGACCTGTGCGCTACCTCTCAAGAAGCTC 1404  
Db 341 ValLeuThrGlnLysLeuGlyProTrpArgProValAlaTyrLeuSerLysLysLeu 360  
Qy 1405 GATCTGTAGCCAGTGGTGGCCCATATGCTGAAGGCTATCGAGCTGTGGCCATCTG 1464  
Db 361 AspProValAlaAlaGlyTrpProProCysLeuArgMetValAlaAlaIleAla 379  
Qy 1465 GTCAGGAGCGCTGACAAATTCGATTTGGCAGAGAATATACTGTAATAGCCCCCATGCA 1524  
Db 380 ThrLysAspAlaGlyLysLeuThrMetGlyGlnProLeuValIleAlaProHisAla 399  
Qy 1525 TTGAGAGACATGTTTCGGCAGCCCCCAGACCGATGATGACCAACGCCCGCATGCCAC 1584  
Db 400 ValGluAlaIleValLysGlnProProAspArgTrpLeuSerAsnAlaArgMetThrHis 419  
Qy 1585 TATCAAGCCTGCTTCTC---ACAGAGGGTACGTTTCGCTCCACGACCGCTCTCAAC 1641  
Db 420 TyrGlnAlaLeuLeuAspThrAspArgValGlnPheGlyProValValAlaLeuAsn 439  
Qy 1642 CTGCGCACTTCTCTCGCTGAAGAGACTGATGAACCAAGTACCTCATGATTCGCACTA 1701

Db 440 ProAlaThrLeuLeuPro---LeuProGluGluGlyLeuGlnHisAsnCysLeuAspIle 458  
Qy 1702 TTGATTGAGGAGACTGGGTCCGCAAGACCTTACA 1737  
Db 459 LeuAlaGluAlaHisGlyThrArgProAspLeuThr 470  
RESULT 9  
US-10-448-871A-30  
; Sequence 30, Application US/10448871A  
; Publication No. US20040093169A1  
; GENERAL INFORMATION:  
; APPLICANT: Mizushina, Yoshiyuki  
; APPLICANT: Sakaguchi, Kengo  
; APPLICANT: Sugawara, Fumio  
; TITLE OF INVENTION: A METHOD FOR DESIGNING A MOLECULAR STRUCTURE OF AN INHIBITOR FOR  
; FILE REFERENCE: 03316/LH  
; CURRENT APPLICATION NUMBER: US/10/448,871A  
; CURRENT FILING DATE: 2003-05-29  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn version 2.0  
; SEQ ID NO 30  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: Radiation murine leukemia virus  
US-10-448-871A-30  
Alignment Scores:  
Pred. No.: 1,17e-140 Length: 471  
Score: 1846.00 Matches: 338  
Percent Similarity: 85.1% Conservative: 63  
Best Local Similarity: 71.8% Mismatches: 68  
Query Match: 29.9% Indels: 2  
DB: 4 Gaps: 2  
US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-10-448-871A-30 (1-471)  
Qy 325 GGTTTGGCAAGCAGTTCCTCCCAAGTATTCACTGAAGGCCAGTGCCACACCACTG 384  
Db 1 GlyLeuAlaValArgGlnAlaProLeuIleLeuProLeuLysAlaThrSerThrProVal 20  
Qy 385 TCAGTCAACAGTACCCCTTCAGTAAAGAACTCAAGAAGGAATTCGCCCGCATGTCCAA 444  
Db 21 SerIleLysGlnTyrProMetSerGlnGluAlaLysLeuGlyIleLysProHisIleGln 40  
Qy 445 AGATTAAATCCCAACAGGGCATCCTAGTTCCTGCTCAATCTCCCTGGAATACTCCCTGCTA 504  
Db 41 ArgLeuLeuAspGlnGlyIleLeuValProCysGlnSerProTrpAsnThrProLeuLeu 60  
Qy 505 CCGGTTAGAAAGCCTGGGACTAATGACTATCGACCAGTACAGGACTTCGAGAGAGTCAAT 564  
Db 61 ProValLysLysProGlyThrAsnAspTyrArgProValGlnGlyLeuArgGluValAsn 80  
Qy 565 AAACGGGTGACAGGATATACACCCCAACAGTCCCGCAACCCCTTATAACCTCTTGTGTCTCTC 624  
Db 81 LysArgValGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerGlyLeu 100  
Qy 625 CCACCCCAACCGAGCTGTATACAGTATGAGCTTAAAGGATGCTTCTTCTGCTGAGA 684  
Db 101 ProThrSerHisArgTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuArg 120  
Qy 685 TTACACCCCACTAGCCCAACCACTTTTTCGCTTCGATCGAGATCCAGGTACGGGAGA 744  
Db 121 LeuHisProThrSerGlnProLeuPheAlaSerGluTrpArgProGlyMetGlyIle 140  
Qy 745 ACCGGCAGCTCACCTGACCGCTGCGCCCAAGGTTCAAGAACTCCCGCCACCATCTTT 804  
Db 141 SerGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPhe 160  
Qy 805 GACGAAGCCCTACACAGAGACTGCGCCAACTTCAGGATCCAAACACCTTCAGGTGACCCCTC 864  
Db 161 AspGluAlaLeuHisArgGlyLeuAlaAspPheArgIleGlnHisProAspLeuIleLeu 180

Qy	865	CTCAGTACGTGGATGACCTGCTCTCTGGCGGAGACCCAAACAGGACTGCTTTAGAAGGC	924
Db	181	LeuGlnTyValAspAspLeuLeuLeuAlaThrSerGluLeuAspCysGlnGlnGly	200
Qy	925	ACGAAGGCACCTACTGCTGGAATTCGTGACCTAGGCTACAGAGCCTCTGCTTAAAGGCC	984
Db	201	ThrArgAlaLeuLeuLysThyLeuGlyAsnLeuGlyTyArgAlaSerAlaLysLysAla	220
Qy	985	CAGATTTCAGAGAGAGAGGTAAACATACTTGGGGTACAGTTTGGCGGACGGCGACCATGG	1044
Db	221	GlnIleCysGlnLysGlnValLysTyLeuGlyTyLeuLeuArgGluGlyGlnArgTyr	240
Qy	1045	CTGAGGGGACGACGAGAAACCTGTGTAGTCAGATACCGGCCCCAACACAGACCAACAA	1104
Db	241	LeuThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGln	260
Qy	1105	ATGAGAGAGTTTGGGGACAGCTGGATTGTGCAGCTGTGCATCCGGGGTTTGGACC	1164
Db	261	LeuArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTyrPheProArgPheAlaGlu	280
Qy	1165	TTACAGCGCCACCTCTACCGCTAACCAAGAAAAAGGGAATTCCTCGGCTCTCTGAG	1224
Db	281	MetAlaAlaProLeuTyProLeuThyLysThrGlyThrLeuPheAsnTyrGlyProAsp	300
Qy	1225	CACCAAGGACGATTGTGATGCTATCAAAAAGCCCTGCTGAGCGCACCTGCTGGCCCTC	1284
Db	301	GlnGlnLysAlaTyHisGluIleLysGlnAlaLeuLeuThrAlaProAlaLeuGlyLeu	320
Qy	1285	CCTGACGTAACTAAACCTTTACCTTTATGTGTGATGAGGTAAAGGATAGCCCGGGGA	1344
Db	321	ProAspLeuThyLysProPheGluLeuPheValAspGluLysGlnGlyTyAlaLysGly	340
Qy	1345	GTTTTAACCCAAACCTTAGGACCATGAGAAAGCCTCTCGCCTACCTGTCAAAGAAGCTC	1404
Db	341	ValLeuThrGlnLysLeuGlyProTyrArgProValAlaTyLeuSerLysLysLeu	360
Qy	1405	GATCCTGTAGCCAGTGTGGCCCATATGCTGAAGCTATCGCAGCTGTGGCCATCTG	1464
Db	361	AspProValAlaAlaGlyTyrProCysLeuArgMetValAlaAlaIleAlaValLeu	380
Qy	1465	GTCAGGACGCTGACAAATTGACTTTGGGACAGAATATATACTGTAATAGCCCCCATGCA	1524
Db	381	ThrLysAspAlaGlyLysLeuThrMetGlyGlnProLeuValIleLeuAlaProHisAla	400
Qy	1525	TTGGAGACATCGTTCCGCGACCCCGACGATGGATGACCAACGCCCGCATGACCCAC	1584
Db	401	ValGluAlaLeuValLysGlnProProAspArgTyrLeuSerAsnAlaArgMetThrHis	420
Qy	1585	TATCAAGCCTGCTCTC---ACAGAGAGGTACAGTTGCTCTCACCGCGCTCTCAAC	1641
Db	421	TyrGlnAlaMetLeuLeuAspThrAspArgValGlnPheGlyProValAlaLeuAsn	440
Qy	1642	CCTGCCACTCTTCTGCTGAAGAGACTGTGAACCAAGTGTGACTCATGATTGCCATCACTA	1701
Db	441	ProAlaThrLeuLeuPro---LeuProGluGluGlyAlaProHisAspCysLeuGluIle	459
Qy	1702	TTGATTGAGAGACTGGGGTCCGCAAGGACCTT	1734
Db	460	LeuAlaGluGluThrGlyThrArgArgAspLeu	470

## RESULT 10

US-10-450-763-42377

US-10-430-703-4237  
; Sequence 42377, Application US/10450763

; Sequence 42377; Application US97000001; Publication No. US20050196754A1

**GENERAL INFORMATION:**

APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

FILE REFERENCE: 790CIP3/US

; CURRENT APPLICATION NUMBER: US/10/450,763

; CURRENT FILING DATE: 2003-06-11

; PRIOR APPLICATION NUMBER: PCT/US01/08631

;  
; PRIOR FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: 09/540,217

```

; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 42377
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1064)..(1107)
; OTHER INFORMATION: CTF/NF-I proteins domain identified by eMATRIX, accession
; OTHER INFORMATION: number BL00349C, p-value=1.000e-09, raw score of 9.33
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1241)..(1394)
; OTHER INFORMATION: Integrase core domain identified by pFam, accession name rve,
; OTHER INFORMATION: E-value=1.4e-33, pFam score of 120.1
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1438)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-42377

```



Db 581 IleProSerProTyThrLeuLeuSerGlnIleProGluGluAlaGluTrpPheThrVal 600  
Qy 652 TTGAGCTTAAAGGATGCTTCTTCGCTGAGATTACACCCACTAGCAACCACTTTT 711  
Db 601 LeuAspLeuLysAspAlaPhePheCysIleProValHisProAspSerGlnPheLeuPhe 620  
Qy 712 GCCTTCGAATGGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACCTGGACCCGACTG 771  
Db 621 AlaPheGlu-----AspPro-----SerAsnProThrSerGlnLeuThrTrpIrrValLeu 637  
Qy 772 CCCAAGGGTTCAAGAACTCCCGACCACTTTTGACGAAGCCCTACACAGAGACTGGCC 831  
Db 638 ProGlnGlyPheArgAspSerProHisLeuPheGlyGlnAlaLeuAlaGlnAspLeuSer 657  
Qy 832 AACTTCAGATCCAAACCCCTCAGGTGACCTCCTCCAGTAGTGAGTGCCTGCTTCTG 891  
Db 658 GlnPhe-----SerTyLeuAspThrProValLeuGlnCysMetAspAspLeuLeuLeu 675  
Qy 892 GCGGGAGCCACCAACAGGACTGCTTAGAAGGCACGAAGGCATCTACTCTGGAATTGTCT 951  
Db 676 AlaAlaArgSerGluThrLeuCysHisGlnAlaThrGlnAlaLeuLeuAsnPheLeuThr 695  
Qy 952 GACCTAGCTACAGAGCCTCTCTGAAGAGCCCAAGATTTCAGAGAGAGGTAAACATAC 1011  
Db 696 ThrCysGlyTyrlsValSerLysProLysAlaGlnLeuCysSerGlnGlnValLysCys 715  
Qy 1012 TTGGGTACAGTTTCGGGAGCGGCGAGCGATGGCTGACGGAGGCACGGAAGAAACTGTA 1071  
Db 716 LeuGlyLeuLysLeuSerLysValThrArgAlaLeuSerGluGluArgIleGlnProfile 735  
Qy 1072 GTCCAGATACCGGCCCAACACCAACCAAAACAAATGAGAGAGTTTTCGGGACAGCTGGA 1131  
Db 736 LeuAlaTyProTyProLysThrLeuLysGlnLeuArgGlyPheLeuGlyIleThrGly 755  
Qy 1132 TTTTGCAGACTGTGATCCCGGGTTTCGCACTTTAGCAGCCCACTACTACCCGCTAAC 1191  
Db 756 PheCysArgIleTrpIleProArgTyrlsGlyLysIleAlaArgProLeuTyThrLeuIle 775  
Qy 1192 AAAGAAAGGGGAATTCCTC-----TGGGCTCTGAGCACACGAGGCA 1236  
Db 776 LysGluThrGlnLysAlaAsnThrHisLeuValArgTrpTrpProGluAlaGluAlaAla 795  
Qy 1237 TTTGATGCTATCAAAAGCCCTGCTGAGCGCACTGCTCTGGCCCTCCCTGACGTAAC 1296  
Db 796 PheHisAlaLeuLysLysAlaLeuMetGlnAlaProValLeuSerLeuLeu-----ThrGly 814  
Qy 1297 AAACCTTTACCTTTATGTGATGAGCGGTAA-----GGAGTA 1335  
Db 815 GlnAspPheSerSerTyValThrLysAsnLysGlnThrLysLysLysThr\*\*\*Ile 834  
Qy 1336 GCGCGGGAGTTTAAACCAACCCCTAGGACCATGGAGAAGACCTGTGCTCCTACTGTCA 1395  
Db 835 AlaLeuArgValLeuAlaLeuVal\*\*\*GlyThrSerLeuGlnProValAlaTyrlsSer 854  
Qy 1396 AAGAGCTCGATCGTGTAGCCAGTGTGGGCCATATGCTGAAGGCTATCGACGCTGT 1455  
Db 855 LysGluIleAspValValAlaLysGlyGlnProHisCysLysLeuArgValValAlaVal 874  
Qy 1456 GCCATACCTGGTCAAGAGCGCTGACAAATTTGACTTTGGGACAGATATAACTGTATAGCC 1515  
Db 875 AlaValLeuValSerGluAlaValLysIleIleGlnGlyArgAsnLeuThrValTrpThr 894  
Qy 1516 CCCCATGATGGAGAACATGCTTCGGGACGCCGCCCGATGATGATGACCAACGCCCGC 1575  
Db 895 SerHisAspValAsnGlyIleLeuThrAlaLysGlyAspLeuTrpLeuSerAspAsnHis 914  
Qy 1576 ATGACCCACTATCAAGCTGCTCTCACA---GAGAGGGTCACGTTCTGCTCCACCGCC 1632  
Db 915 LeuLeuLysTyrlsGlnAlaLeuLeuGluGlyProMetLeuArgLeuCysThrCysAla 934  
Qy 1633 GCTCTCAACCCCTGCCACTCTTCTGCTGAAGAGACTGATGAACCAAGTGCATCATGTC 1692  
Db 935 AlaLeuAsnLeuAspThrPheLeuPro---HisAsnGluGluLysIleGluHisAsnCys 953

Qy 1693 CATCAACTATTGATTGAGGAGACTGGGGTCCGAAGGACCTTTACAGACATACCCGCTGACT 1752  
Db 954 GlnGlnValIleAlaGlnThrTyAlaThrArgGlyAspHisLeuGluValProLeuThr 973  
Qy 1753 GGAGAAGTGTCTAACTGGTTTCACCTGACGAGGAGCAGCTATGTGTGGAGGTAAAGAGATG 1812  
Db 974 AspProAsnProAsnLeuTyThrAspGlyArgSerPheValGluLysGlyLeuGlnLys 993  
Qy 1813 GCTGGGGGGGGTGTGGAGCCGACCCGACGATCTGGGGCAGCAGCTCCCGGAAGGA 1872  
Db 994 ValGlyTyrlsAlaValSerAspAsnGlyIleLeuGluSerAsnProLeuThrProGly 1013  
Qy 1873 ACTTCAGCACAAAAGGCTGAGCTCATGCCCTCAGCAAGCTTTTGGCTGGCGCGAAGGG 1932  
Db 1014 ThrSerAlaGlnLeuAlaLys-----LeuThrArgAlaLeuGluLeuGlyGly 1030  
Qy 1933 AAATCCATAAACATTTATACGACGACGAGTATGCTTTTGCAGCTGCACACGTACATGG 1992  
Db 1031 LysArgValAsnIleTyThrAspSerLysTyrlsAlaTyLeuValLeuHisAlaHisAla 1050  
Qy 1993 GCCATCTATAAAACAAAGGGGTTCCTTACCTCAGCAGGAGGAGAAATAAAGAACAAAGAG 2052  
Db 1051 ValIleTrpArgGluArgGluPheLeuThrSerGluGlyThrProIleLysHisGlnGlu 1070  
Qy 2053 GAATTTCTAAGCTATTAGAACCGCTACATTTACCAAAAGCGTAGCTATTATATACACTGT 2112  
Db 1071 AlaIleArgGluLeuLeuAlaValGlnLysProLysGluValAlaValLeuHisCys 1090  
Qy 2113 CCTGCAGCATCAGAAAGCTTAAAGATCTCATATCCAGAGGAAACACGATGGCTGACCGGTT 2172  
Db 1091 ArgGlyHisGlnLysGlyLysGluArgGluIleGluArgAsnArgGlnAlaAspIleGlu 1110  
Qy 2173 GCCAAGCAGGAGCC---CAGGGGTGTTAACTTCTGCTTATAATAAGAAATGCCAAAGCC 2229  
Db 1111 AlaLysArgAlaAlaArgGlnAspProLeuGluMetLeuIleGluGlyProLeuVal 1130  
Qy 2230 -----CCA-----GAACCCAGACGACAGTACACCTAGAA-----GACTGGCAA 2268  
Db 1131 TrpGlyAsnProLeuTrpGluThrLysProGlnTyrlsGluGluIleGluTrp--- 1149  
Qy 2269 GAGATAAAAGATAGACCACTTCTCTGAGACTCCGGAAGGGACCTGCTATACCTCAGAT 2328  
Db 1150 GlyIleSerArgGlyHisSerPhe-----LeuProSerGlyTrpGlnAlaThrGluGln 1167  
Qy 2329 GGGAAGGAATCTCGCCCCCAAGAGAGGTAGATATATGTCACAGATACATCGTCTA 2388  
Db 1168 GlyLysIleProLeuProAlaAlaAsnGlnTrpLysLeuLeuLysThrHisGlnAsn 1187  
Qy 2389 ACCCACTTA-----GGAACTAAACACCTGACGAGTGTGGTCAACATGTGTGCC 2436  
Db 1188 PheHisLeuCysIleAspSerThrHisGlnMetAlaLysLeuLeuPheThrGlyProGly 1207  
Qy 2437 ---TATCATGTTCTGAGGCTACAGGAGTGGCTGACTCGGTGGTCAACATGTGTGCC 2493  
Db 1208 LeuPheLysThrLeuLys-----GlnIleValGlyAlaCysGluVal 1221  
Qy 2494 TGCAGCTGGTTAATGCTAATCCT---TCCAGATGCTCCAGGAAGACAGATAGGGGA 2550  
Db 1222 CysGlnIleAsnAsnProLeuProTyrlsHisGlnAlaProProGlyGluGlnArgThrGly 1241  
Qy 2551 AGCCACCCAGCGCTCACTGGGAAGTGGACTTCACTGAGGTAAAGCCGCTAAATACGGA 2610  
Db 1242 HistySerGlyGluAspTrpGlnLeuAspPheThrHisMet---ProLysSerGlnGly 1260  
Qy 2611 AACAAATACCTATTGTTTTGTAGACACCTTTTTCAGGATGGGTAGAGGCTTATCCTACT 2670  
Db 1261 PheGlnTyrlsLeuLeuValTrpValAspIlePheThrGlyTrpAlaGluAlaPheProCys 1280  
Qy 2671 AAGAAAGAGACTTCAACCGTGTGCTTAAAAATACTGGAGAGAAATTTTCCAGATT 2730  
Db 1281 ArgThrGluLysAspGlnAspValIleLysAlaLeuIleHisGluIleProArgPhe 1300

101	SerProAspArgThrTrpTyrThrValLeuAspLeuLysAspAlaPhePheCysLeuPro	120
685	TTACACCCCACTAGCCAAACACATTTTTCCTTCGAAATGGAGAGATCCAGTACGGGAGAGA	744
121	LeuAlaProGlnSerGlnGluLeuPheAlaPheGluTrpArgAspProGluAr-ggGlyLe	140
745	ACGGGGCAGCTCACCTGGACCCGACTGCCCCAAAGGGTTCAAGAACTCCCGACCATCTTT	804
141	SerGlyGlnLeuThrTrpArgLeuProGlnGlyPheLysAsnSerProThrLeuPhe	160
805	GACGAAGCCCTACACAGAGACCTGGCCCACTTCAGATCCAAACACCTCAGGTGACCTC	864
161	AspGluAlaLeuHisArgAspLeuThrAspPheArgThrGlnHisProGluValThrLeu	180
865	CTCCAGTACCTGGATGACCTGCTTCTGGGGGAGCCACCAACAGGACTGCTTTAGAAGGC	924
181	LeuGlnTyrValAspAspLeuLeuAlaProThrLysGluAlaCysIleAr-ggGly	200
925	ACGAAGCCTACTCTGCTGGAAATTGCTCGACTAGGCTACAGACCTCTGCTAAGAAGGCC	984
201	ThrLysHisLeuLeuArgGluLeuGlyAspLysGlyTyrArgAlaSerAlaLysLysAla	220
985	CAGATTTCAGGAGAGAGGTAAACATCTTGGGGTACAGATTGGCGGACGGCAGCATGG	1044
221	GlnIleCysGlnLysValThrTyrLeuGlyTyrIleLeuSerGluGlyLysAr-gtrp	240
1045	CTGACGGAGCCAGGAAGAAACTGTAGTCCAGATACCGGCCCCCAACACAGCCCAACAA	1104
241	LeuThrProGlyArgIleGluThrValAlaHisIleProProGlnAsnProAr-ggGlu	260
1105	ATGAGAGAGTTTTCGGGACAGCTGGATTTTGACACTGTGATCCCGGGTTTGGCACC	1164
261	ValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGlu	280
1165	TTAGACGCCCACTCTACCGGCTAACCAAGAAAAAGGGGAATTCTCTGGGCTCCTGAG	1224
281	LeuAlaAlaProLeuTyrAlaLeuThrLysGluSerAlaProPheThrTrpGlnGluLys	300
1225	CACCAGAAGCATTTGCATGCTATCAAAAGGCCCTCTCAGCGCACCTGCTCGGCCCTC	1284
301	HisGlnSerAlaPheGluAlaLeuLysGluAlaLeuLeuSerAlaProAlaLeuGlyLeu	320
1285	CTGACGTAATAACCCCTTTTACCCTTTATGTGGATGAGCGTAAGGAGTAGCCCGGGGA	1344
321	ProAspThrSerLysProPheThrLeuPheIleAspGluLysGlnGlyIleAlaLysGly	340
1345	GTTTTAAACCAACCCCTAGACCATGAGAGACCTGTGCGCTACTCTGTCAAAGAAGCTC	1404
341	ValLeuThrGlnLysLeuGlyProTrpLysArgProValAlaTyrLeuSerLysLysLeu	360
1405	GATCCTGTGACCATGGTTGGCCCATATCGCTCAAGGCTATCGCAGCTGTGGCCATACG	1464
361	AspProValAlaAlaGlyTrpProProCysLeuArgIleMetAlaAlaThrAlaMetLeu	380
1465	GTCAAGGACGCTGACAAATTTGACTTTGGACAGAAATATACTGTATAGCCCCCATGCA	1524
381	ValLysAspSerAlaLysLeuThrLeuGlyGlnProLeuThrValIleThrProHisAla	400
1525	TTGGAGAACATGTCGCGAGCCGCCAGACCGATGGATGACCAAC---GCCCGCATGACC	1581
401	LeuAlaAlaIleValArgGlnThrProAspArgTrpIleThrAsnAlaAlaArgLeuThr	420
1582	CACTATCAAGCTGCTCTC---ACAGAGAGGCTCACGTTTCGCTCCACAGCCGCTC	1638
421	HisTyrGlnAlaLeuLeuLeuAspThrAspArgIleGlnPheGlyProProValThrLeu	440
1639	AACCTCGCCACTCTTCTGCTGAAGACTGTATGAACCATGACTCATGATTTGCCATCAA	1698
441	AsnProAlaThrLeuLeuProAlaProGluAspGlnGlnSerAlaHisAspCysArgGln	460
1699	CTATTGATTGAGAGACTGGGTCGCCAGGACCTTT	1734
461	ValLeuAlaGluThrHisGlyThrArgGluAspLeu	472

## RESULT 12

US-10-448-871A-26  
; Sequence 26, Application US/10448871A  
; Publication No. US20040093169A1  
; GENERAL INFORMATION:  
; APPLICANT: Mizushiba, Yoshiyuki  
; APPLICANT: Sakaguchi, Kengo  
; APPLICANT: Sugawara, Fumio  
; TITLE OF INVENTION: A METHOD FOR DESIGNING A MOLECULAR STRUCTURE OF AN INHIBITOR FOR  
; FILE REFERENCE: 03316/LH  
; CURRENT APPLICATION NUMBER: US/10/448,871A  
; CURRENT FILING DATE: 2003-05-29  
; NUMBER OF SEQ ID NOS: 59  
; SEQ ID NO 26  
; LENGTH: 472  
; TYPE: PRT  
; ORGANISM: Baboon endogenous virus  
US-10-448-871A-26

Alignment Scores:  
Pred. No.: 2,33e-134 Length: 472  
Score: 1768.50 Matches: 328  
Percent Similarity: 81.3% Conservative: 55  
Best Local Similarity: 69.6% Mismatches: 87  
Query Match: 28.6% Indels: 1  
DB: 4 Gaps: 1

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-10-448-871A-26 (1-472)

QY	325	GGTTGGCAAGCAAGTTCCTCCCAACAGTTATTCACTGAAGGCCAGTGCACACCAAGTG	384
DB	1	GlyArgAlaLysCysGlnAlaProIleIleAspLeuLysProThrAlaValProVal	20
QY	385	TCAGTCACACAGTACCCCTTCAGTAAGAAGCTCAAGAAGAAATTCGGCCCGCATGTCCAA	444
DB	21	SerIleLysGlnTyrProMetSerLeuGluAlaHisMetGlyIleArgGlnHisIleIle	40
QY	445	AGATTAATCAACACAGGCGATCTAGTTCCTGTCCATCTCCCTGGAATPACTCCCTGCTA	504
DB	41	LysPheLeuGluLeuGlyValLeuArgProCysArgSerProTrpAsnThrProLeuLeu	60
QY	505	CGGTTAGAAAGCTGGGACTAATGACTATCGACAGTACAGGACTTCAGAGAGGTCAAT	564
DB	61	ProValLysLysProGlyThrGlnAspTyrArgProValGlnAspLeuArgGluIleAsn	80
QY	565	AAACGGGTGCAGGATATACACCAACAGTCCCGAACCTTATACCTCTCTGTGTGCTCTC	624
DB	81	LysArgThrValAspIleHisProThrValProAsnProTyrAsnLeuLeuSerThrLeu	100
QY	625	CCACCCCAACCGAGCTGTGTATACAGTATGGACTTAAAGGATGCCCTTCTTCTGCTCGAGA	684
DB	101	LysProAspTyrSerTrpTyrThrValLeuAspLeuLysAspAlaPheCysLeuPro	120
QY	685	TTACACCCCACTAGCCAAACCACTTTTTCCTTCGATCGAGAGATCCAGGTACGGGAAGA	744
DB	121	LeuAlaProGlnSerGlnGluLeuPheAlaPheGluTrpLysAspProGluArgGlyIle	140
QY	745	ACCGGGCAGCTCAGCTGACCCGAGTCCCAAGGGTTCAAGACTCCCGCCAGCATCTTT	804
DB	141	SerGlyGlnLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPhe	160
QY	805	GACCAAGCCCTACACAGAGACTGCGCAACTTCAGGATCCAAACCCCTCAGGTGACCCCTC	864
DB	161	AspGluAlaLeuHisArgAspLeuThrAspPheArgThrGlnHisProGluValThrLeu	180
QY	865	CTCCAGTACGTGGATGACTGCTGCTTCTGGGGAGCCACCAACAGGACTGCTTAGAGGC	924
DB	181	LeuGlnTyrValAspAspLeuLeuAlaProThrLysLysAlaCysThrGlnGly	200
QY	925	ACGAAGGCACACTGCTGGAATTGCTGACCTTAGGCTACAGAGCCTCTGCTAAGAAGGCC	984

DB	201	ThrArgHisLeuLeuGlnGluLeuGlyGluLysGlyTyrArgAlaSerAlaLysLysAla	220
QY	985	CAGATTTGCAGGAGAGAGGTAAACATCTTGGGTACAGTTTGGGGGCGGCGGCGATGG	1044
DB	221	GlnIleCysGlnThrLysValThrTyrLeuGlyTyrIleLeuSerGluGlyLysArgTrp	240
QY	1045	CTGACGGAGGACACGGAAGAAACGTAGTCAGATACCGGCCCAACACACAGCCAAACAA	1104
DB	241	LeuThrProGlyArgIleGluThrValAlaArgIleProProAsnProArgGlu	260
QY	1105	ATGAGAGAGTTTTTGGGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTCGAC	1164
DB	261	ValArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGlu	280
QY	1165	TTAGCAGCCCTCTCTACCCGCTAACCAAGAAAGGGGAATTCCTCTGGGCTCTTGAG	1224
DB	281	LeuAlaAlaProLeuTyrAlaLeuThrLysGluSerThrProPheThrTrpGlnThrGlu	300
QY	1225	CACCAGAAGCANTTGTATCTATCAAAAGCCCTGTGAGCGACCTGTCTGCGCCCTC	1284
DB	301	HisGlnLeuAlaPheGluAlaLeuLysLysAlaLeuLeuSerAlaProAlaLeuGlyLeu	320
QY	1285	CCTCAGCTAACTAAACCTTTTACCCTTTTATGTGTGAGCGGTAAAGGAGTAGCCCGGA	1344
DB	321	ProAspThrSerLysProPheThrLeuPheLeuAspGluArgGlnGlyIleAlaLysGly	340
QY	1345	GTTTTAACCAACCCCTAGGACCATGAGAGAGACCTGTGCGCTACCTGTCAAAGAGCTC	1404
DB	341	ValLeuThrGlnLysLeuGlyProTrpLysArgProValAlaTyrLeuSerLysLysLeu	360
QY	1405	GATCCTGTAGCCAGTGTGTCCTCATATGCTGAAGCTATCGCAGCTGTGCGCCATCTG	1464
DB	361	AspProValAlaAlaGlyTrpProCysLeuArgIleMetAlaAlaThrAlaMetLeu	380
QY	1465	GTCAAGGACGCTGCAAAATTCAGCTTTGGGACAGAAATATACTGTAAATAGCCCCCATG	1524
DB	381	ValLysAspSerAlaLysLeuThrLeuGlyGlnProLeuThrValIleThrProHisThr	400
QY	1525	TTGAGAACATCTGTCGCGAGCCCCCAGACCGATGGATGACCAACGCGCGATGACCCAC	1584
DB	401	LeuGluAlaIleValArgGlnProProAspArgTrpIleThrAsnAlaArgLeuThrHis	420
QY	1585	TATCAAGACCTGCTTCTC---ACAGAGAGGTGACGTTTCGCTCCACACGCGCTCTCAAC	1641
DB	421	TyrGlnAlaLeuLeuLeuAspThrAspArgValGlnPheGlyProProValThrLeuAsn	440
QY	1642	CCTGCCACTCTTCTGCTGGAAGAGACTGATGAACACAGTGACTCATGATTGCCATCAACTA	1701
DB	441	ProAlaThrLeuLeuGluValProGluAsnGlnProSerProHisAspCysArgGlnVal	460
QY	1702	TTGATTGAGGAGACTGGGTCCCGCAAGACCTT	1734
DB	461	LeuAlaGluThrHisGlyThrArgGluAspLeu	471

## RESULT 13

US-10-450-763-33781  
; Sequence 33781, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom

```

; SEQ ID NO 33781
; LENGTH: 1267
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1066)..(1109)
; OTHER INFORMATION: CTF/NF-I proteins domain identified by eMATRIX, accession
; OTHER INFORMATION: number BL00349C, p-value=1.000e-09, raw score of 9.33
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (975)..(1120)
; OTHER INFORMATION: RNase H domain identified by Pfam, accession name rnaseH, E-
; OTHER INFORMATION: value=3.6e-34, Pfam score of 126.9
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1267)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-33781

Alignment Scores:
Pred. No.:      8.54e-107      Length:      1267
Score:          1431.50      Matches:      347
Percent Similarity: 55.6%      Conservative: 136
Best Local Similarity: 39.9%      Mismatches:   315
Query Match:    23.2%      Indels:       71
DB:              5          Gaps:         22

US-10-723-552-3_COPY_2307_5741 (1-3435) x US-10-450-763-33781 (1-1267)

QY 52 GACTTGGAGTGGACGGGTAAACCACTCGTTTCTGTCATACCTGATGCCAGCACCC 111
Db 429 AspTrpGlyAlaLeuProPheSerHisAlaPheLeuIleMetProGluSerLeuThrPro 448
QY 112 CTCCTAGGTAGACTTATTGACCAAGATGGAGCACAAATTTCTTTTGAACAGGAA 171
Db 449 LeuLeuGluArgGluIleLeuValLysAlaGlyAlaIleHisLeuAsnIleGlyGlu 468
QY 172 CCAGAAGTGTCTGCAAAATAACAACTATCATCTGTGTGACCTCCCAATTAGATGACGAA 231
Db 469 -----GlyThrProIleCysArgLeuLeuPhe----- 477
QY 232 TATCGACTATACTCTCCCTAGTAAAGCTGTATCAAAATATACAAATTCGTGTGGAACAG 291
Db 478 -----GluGluGlyIleSer----- 482
QY 292 TTTTCCCAAGCTGGCAGAAACCGCAGGATGGTTTGGCAAGCAAGTTCCCCCAAA 351
Db 483 ---ProGluValTrpAlaThrGluGlyGlnTyrGlyGlnAlaLysAsnAlaHisPheVal 501
QY 352 GTTATTCAACTGAAGGCCAGTGCCACACACAGTGTGTCAGTACAGACAGTACCCCTTGAGTAA 411
Db 502 GlnValLysLeuLysAspSerThrSerPheProTyrGlnArgGlnTyrProLeuArgPro 521
QY 412 GAAGCTCAAGAGGAATTCGGCCGCATGTCCAAAGATTAATCCAAAGGCATCTTAGTT 471
Db 522 GluAlaGlnArgLeuGlnLysIleValLysAspLeuLysAlaGlnGlyLeuValLys 541
QY 472 CCTGTCCAATCTCCCTGGAATATCTCCCTGCTACCGGTAGAAAGCCTGGGACTAATGAC 531
Db 542 ProTyrSerProCysAsnThrProIleLeuGlyValGlnLysPro---LysArgGln 560
QY 532 TATCGACAGTACAGGACTTGAGAGAGGTCAATAAACGGGTGAGGATATACACCCCAACA 591
Db 561 TrpArgLeuValGlnAspLeuArgIleIleAsnGluAlaValPheProLeuTyrProAla 580
QY 592 GTCCGAAACCTTATAACCTCTGTGTGCTCTCCACCCCAACAGGAGCTGGTATACAGTA 651
Db 581 IleProSerProTyrThrLeuLeuSerGlnIleProGluGluAlaGluTrpPheThrVal 600
QY 652 TTGAGCTTAAGGATGCTCTCTGCTGCTGAGATTACACCCACTACCCCAACACTTTT 711
Db 601 LeuAspLeuLysAspAlaPhePheCysIleProValHisProAspSerGlnPheLeuPhe 620

```

```

QY 712 GCCTTCGAATGGAGAGATCCAGGTACCGGAAGAAACCGGCGAGCTCACCTGACCCGACTG 771
Db 621 AlaPheGlu-----AspPro---SerAsnProThrSerGlnLeuThrTrpValLeu 637
QY 772 CCCCAAGGGTTCAAGAACTCCCGACCATCTTTGACGAGAGCCCTACACAGAGACCTGGCC 831
Db 638 ProGlnGlyPheArgAspSerProHisLeuPheGlyGlnAlaLeuAlaLeuAlaLeuSer 657
QY 832 AACTTTCAGGATCCAAACACCCCTCAGGTGACCTCTCCAGTACGTGCGTGGATGACCTGCTT 891
Db 658 GlnPhe-----SerTyrLeuAspThrProValLeuGlnCysMetAspAspLeuLeu 675
QY 892 GCGGAGCCACCAACAGGACTGCTTAGAAGGACGAGGACACTACTGCTGGAATTTGCT 951
Db 676 AlaAlaArgSerGluThrLeuCysHisGlnAlaThrGlnAlaLeuLeuAsnPheLeuThr 695
QY 952 GACCTAGGCTACAGAGCTCTGTGTAAGAGGCCAGATTTGACGAGAGAGGTAACTATAC 1011
Db 696 ThrCysGlyTyrLysValSerLysProLysAlaGlnLeuCysSerGlnGlnValLysCys 715
QY 1012 TTGGGGTACAGTTTTCGCGGACGGGCAGCGATGGCTGACGGAGGACCGAAGAAACTGTA 1071
Db 716 LeuGlyLeuLysLeuSerLysValThrArgAlaLeuSerGluGluArgIleGlnProIle 735
QY 1072 GTCAGATACCGGCCCAACACACAGCCAAACAAATGACGAGAGAGTGTGTTGGGACAGCTGGA 1131
Db 736 LeuAlaTyrProTyrProLysThrLeuLysGlnLeuArgGlyPheLeuGlyIleThrGly 755
QY 1132 TTTTGCAGACTGTGGATCCCGGGTTTGGCAGCTTAGCAGCCCACTCTACCCGCTAAC 1191
Db 756 PheCysArgIleTrpIleProArgTyrGlyLysIleAlaArgProLeuTyrThrLeuIle 775
QY 1192 AAGAAAAAGGGGAATTCCTCC-----TGGGCTCCTGAGCAGCAGCAGGCA 1236
Db 776 LysGluThrGlnLysAlaAsnThrHisLeuValArgTrpThrProGluAlaGluAla 795
QY 1237 TTTGATGCTATCAAAAGGCGCTGCTGAGCGCACCTGCTGCGCCCTCCCTGACGTAACT 1296
Db 796 PheHisAlaLeuLysLysAlaLeuMetGlnAlaProValLeuSerLeuLeu---ThrGly 814
QY 1297 AAACCTTTTACCCTTTATGTGGATGACGCTAAG-----GGAGTA 1335
Db 815 GlnAspPheSerTyrValThrLysAsnLysGlnThrLysLysLysLysLysLysLysLys 834
QY 1336 GCCCGGGGAGCTTTTAAACCCAAACCTAGGACCATGGAGAGAGACTGTGCGCTACCTGCA 1395
Db 835 AlaLeuArgValLeuAlaLeuVal**GlyThrSerLeuGlnProValAlaTyrLeuSer 854
QY 1396 AAGAAGCTCCTGATCCTGTAGCCAGTGGTGGCCCATATGCTGGAAGGCTATCGCAGCTGTG 1455
Db 855 LysLysThrAspValValAlaLysGlyTrpProHisCysLeuTrpValMetAlaAlaIle 874
QY 1456 GCATACCTGTCAAGGACGCTGACAAATGACTTTGGGACAGAAATATACTGTAATAGCC 1515
Db 875 AlaValLeuIleSerLysAlaValLysMetIleGln**ArgAspLeuThrValTrpThr 894
QY 1516 CCCCATGCACTGGAGAACATCGTTTCGGCAGCCCCCAGACCGATGATGACCAACGCCCGC 1575
Db 895 SerHisAspValAsnGlyIleLeuThrAlaLysGlyAspLeuTrpLeuSerAspAsnCys 914
QY 1576 ATGACCCACTATCAAAAGCCTGCTTCTCACAGAGAGGTC---ACGTTTCGCTCCACCAAGCC 1632
Db 915 LeuLeuLysCysGlnAlaLeuLeuLeuGluGlyProValLeuArgLeuCysThrCysAla 934
QY 1633 GCTCTCAACCTGCCACTCTTCTGCTGAGAGAGACTGATGAACAGTACATCATGATGTC 1692
Db 935 ThrLeuAsnProAlaThrPheLeuPro---AspAsnGluGluLysIleLysHisAsnCys 953
QY 1693 CATCACTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACT 1752
Db 954 GlnGlnValIleSerGlnThrTyrAlaThrArgLysAspLeuLeuGluValProLeuThr 973

```



Db 151 GlnValLysLeuLysAspSerAlaSerPheProTyrGlnArgLysTyrProLeuArgPro 170  
QY 412 GAAGCTCAAGAGGAATTCGGCGCATGTCCTCAAGATTAATCAACAGGGCATCTAGTT 471  
Db 171 GluAlaLeuGlnGly\*\*GlnLysLeuValLysAspLeuLysAlaGlnGlyLeuValLys 190  
QY 472 CCTGTCCAATCTCCCTCGAATACTCCCTGTGTACCGGTAGAGAGCTGGGACTAATGAC 531  
Db 191 ProCysSerProCysAsnThrProLeuGlnGlyValArgLysPro---AsnGlyGln 209  
QY 532 TATCGACCATGACGAGCTTGAGAGAGTCAATAAAGCGGTGACGAGATATACACCAACA 591  
Db 210 TrpArgLeuValGlnAspLeuArgLysLeuAlaValPheProLeuTyrProAla 229  
QY 592 GTCCGGAACCTTATACTCTGTGTGTCTCTCCACCCCAACAGGAGCTGTGTATACAGTA 651  
Db 230 ValSerSerProTyrThrLeuLeuSerLeuLeuProGluGluAlaGluTrpPheThrVal 249  
QY 652 TTGACTTAAAGGATGCTCTCTGCTGAGATTACACCCACTGACCAACCACTTTT 711  
Db 250 LeuAspLeuLysAspAlaPheCysIleProValArgProAspSerGlnPheLeuPhe 269  
QY 712 GCCTTTCGAATGAGAGATCCAGGTACGGGAAGAACCGGGCAGCTCACTGACCCGACTG 771  
Db 270 AlaPheGlu-----AspPro---LeuAsnProThrSerGlnLeuThrTrpThrValLeu 286  
QY 772 CCCAAGGTTCAAGACTCCCGACCATCTTTGACGAAGCCCTACACAGAGACTGGCC 831  
Db 287 ProGlnGlyPheArgAspSerProHisLeuPheGlyGlnAlaLeuAlaGlnAspLeuSer 306  
QY 832 AACTTCAGGATCCAAACCCCTCAGGTGACCTCTCTCAGTACGTGGAGTACCTGCTCTG 891  
Db 307 GlnPhe-----SerTyrLeuAspThrLeuValLeuGlnTyrValAspAspLeuLeu 324  
QY 892 GCGGGAGCCACCAACAGGACTGCTTAGAGCGCAGAGGCACTACTGCTGGAATGCT 951  
Db 325 ValAlaArgSerGluThrLeuCysHisGlnAlaThrGlnGluLeuLeuThrPheLeuThr 344  
QY 952 GACTAGGCTACAGAGCTCTGTGTAAGAGCCAGATTTGACGAGAGAGGTAAACATAC 1011  
Db 345 ThrCysGlyTyrLysValSerLysProLysAlaArgLeuLysCysSerGlnGluLeuArgTyr 364  
QY 1012 TTGGGTACAGTTTGGGGAGCGGAGGGTGGCTGACGGAGGACCGAAGAACTGTA 1071  
Db 365 LeuGlyLeuLysLeuSerLysGlyThrArgAlaLeuSerGluGluArgGlnGlnProIle 384  
QY 1072 GTCCAGATACCGGCCCAACACACAGCAAAACAAATGAGAGAGTTTGGGGACAGCTGGA 1131  
Db 385 LeuAlaTyrProHisProLysThrLeuLysGlnLeuArgGlyPheLeuGlyLeuThrGly 404  
QY 1132 TTTTGCAGACTGTGGATCCCGGGGTTTGGCAGCTTAGCAGCCCACTCTACCCGCTAAC 1191  
Db 405 PheCysArgLysGlnIleProArgTyrThrProIleAlaArgProLeuTyrThrLeuIle 424  
QY 1192 AAGAAAAGGGGAATTCCTCC-----TGGGCTCTGAGCACCAGAGGCA 1236  
Db 425 ArgGluThrGlnLysAlaAsnThrTyrLeuValArgTrpThrProThr---GluValAla 443  
QY 1237 TTTGATCTATCAAAAGGCCCTGCTGAGCGCACTCTCTGCGCCCTCCCTGAGCTAACT 1296  
Db 444 PheGlnAlaLeuLysLysAlaLeuThrGlnAlaProValPheSerLeuPro---ThrGly 462  
QY 1297 AAACCTTTTACCCTTTATGTGGATGAGCGTAAGGAGTAGCCGGGAGTTTAAACCAA 1356  
Db 463 GlnAspPheSerLeuTyrAlaThrGlnLysThrGlyIleAlaLeuGlyValLeuThrGln 482  
QY 1357 ACCCTAGGACCATGAGAGACCTGTGCGCTCACTGTCAAGAGAGCTCGATCTCTGAGCC 1416  
Db 483 ValSerGlyMetSerLeuGlnProValValTyrLeuSerLysGlnIleAspValValAla 502  
QY 1417 AGTGGTTGGCCCATATCCCTGAAGGCTATCGAGCTGTGGCCATCTGTGTCAAGGAGCT 1476  
Db 503 LysGlyTrpProHisCysLeuTrpValMetAlaAlaValAlaValLeuValSerGluAla 522

QY 1477 GACAAATTGACTTTTGGGACAGATATAACTGTAAATAGCCCCCATGTCATTTGGAGAACATC 1536  
Db 523 VallylIleIleGlnGlyArgAspLeuThrValTrpThrSerHisAspValAsnGlyIle 542  
QY 1537 GTTCGGCAGCCCCCAGAGCCGATGATGACCAACCGCCGCATGACCCACTATCAAGGCTG 1596  
Db 543 LeuThrAlaLysGlyAspLeuTrpLeuSerAspAsnHisLeuLeuAsnTyrGlnAlaLeu 562  
QY 1597 CTTCTCAGACAGAGGGTC---ACGTTGCTCCACAGCGCTCTCAACCTCGCCACTCTT 1653  
Db 563 LeuLeuGluGluProValLeuArgLeuArgThrCysAlaThrLeuLysProAlaThrPhe 582  
QY 1654 CTGCTGAAGAGACTGATGAACACGAGTGAATGATGCCATCACTATTAATTGAGGAG 1713  
Db 583 LeuPro---AspAsnGluGluLysIleGluHisAsnCysGlnGlnValIleAlaGlnThr 601  
QY 1714 ACTGGGTCGCGCAGGACCTTACAGACATACCGCTGACTGGAGAGTGTCAACCTGGTTC 1773  
Db 602 TyrAlaAlaArgGlyAspLeuLeuGluValProLeuThrAspProAspLeuAsnLeuTyr 621  
QY 1774 ACTCAGCAAGCAGCTATGTGCGAAGTAAGAGGATGGCTGGCGCGCTGTGGAC 1833  
Db 622 ThrAspGlySerSerLeuAlaGluLysGlyLeuArgLysAlaGlyTyrAlaValIleSer 641  
QY 1834 GGGACCCGCGACGATCTGGCGCAGCGCTGCCGGAAGGAACTTCAGCACAAAAGCTGAG 1893  
Db 642 AspAsnGlyIleLeuGluSerAsnArgLeuThrProGlyThrSerAlaHisLeuAlaGlu 661  
QY 1894 CTCATGCGCTCAGCAAGCTTTGCGGCTGGCCGAGGAAATCCATAACATTTATACG 1953  
Db 662 LeuIleAlaLeuThrTrpAlaLeuGluLeuGlyGlyArgValAsnIleTyrSer 681  
QY 1954 GACAGAGGTATGCTTTGCGACTGCGACACTATCGGCGCATCTATAACAAGGGGG 2013  
Db 682 AspSerLysTyrAlaTyrLeuValLeuHisAlaHisAlaIleTrpArgGluArgGlu 701  
QY 2014 TTGCTTACCTCAGCAGGAGGAAATAAGAACAAAGAGGAAATTTAAGCTTATTGAA 2073  
Db 702 PheLeuThrSerGluGlyThrProIleAsnHisGlnGluAlaIleArgArgLeuLeu 721  
QY 2074 GCGGTACATTTACAAAAGGCTAGCTATTATACACTGTCTGTCATCAGAAAGCTAAA 2133  
Db 722 AlaValGlnLysProLysGluValAlaValLeuHisCysGlnGlyHisGlnGluGlu 741  
QY 2134 GATCTCATATCCAGAGGAAACCATGATGCGTGCACCGGTTGCCAGCAGGAGCCAG 2190  
Db 742 GluArgGluIleGluGlyAsnArgGlnAlaAspIleGluAlaLysLysAlaAlaArg 760

## RESULT 15

US-10-114-104-89  
; Sequence 89, Application US/10114104  
; Publication No. US20030198647A1

## GENERAL INFORMATION:

APPLICANT: PERRON, HERVE  
BESEME, FREDERIC  
BEDIN, FREDERIC  
PARAMOS-BACCALA, GLAUCIA  
KOMURIAN-PRADEL, FLORENCE  
JOLIVET-REYNAUD, COLETTE  
MANDRAND, BERNARD  
GARSON, JEREMY  
TUIKE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHY  
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC

STREET: P.O. BOX 19928

CITY: ALEXANDRIA

STATE: VA

COUNTRY: USA







**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 16:19:01 ; Search time 3.01216 Seconds  
(without alignments)  
2992.933 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_2307\_5741

Perfect score: 6183

Sequence: 1 ATGGTGCCACAGGCAACA.....CTGTCAATAACCTCTCAGAC 3435

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 194028

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSSWEB.spool/US10723552/runat.14022006.125151.13368/app.query.fasta\_1  
-DB=Published Applications AA New -QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pt0 -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs02p  
-USPR=US10723552 @CGN.1.1.17 @runat.14022006.125151.13368 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA New:

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pdb:
- 2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pdb:
- 3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pdb:
- 4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pdb:
- 5: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pdb:
- 6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pdb:
- 7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pdb:
- 8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pdb:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1660.5	26.9	1189	6	US-10-821-234-1209
2	1612	26.1	1150	6	US-10-821-234-1083
3	431	7.0	912	7	US-11-042-988-12
4	301	4.9	562	6	US-10-507-928-10
5	301	4.9	562	6	US-10-507-928-12
6	301	4.9	562	7	US-11-029-465-10
7	287.5	4.6	561	7	US-11-029-465-12
8	217.5	3.5	210	7	US-11-022-562-221
9	210.5	3.4	422	7	US-11-230-251-26

10	174.5	2.8	188	7	US-11-234-786-592
11	146.5	2.4	14130	7	US-11-175-689-9
12	143.5	2.3	16990	7	US-11-175-689-7
13	126.5	2.0	5712	7	US-11-143-980-47
14	126	2.0	745	6	US-10-985-561-559
15	126	2.0	745	7	US-11-109-156-14
16	126	2.0	745	7	US-11-222-158-4
17	124.5	2.0	937	7	US-11-017-550-66
18	121.5	2.0	2410	7	US-11-175-689-8
19	121.5	2.0	7968	7	US-11-186-731-5
20	121	2.0	900	7	US-11-182-016-37
21	121	2.0	2591	6	US-10-453-372-718
22	121	2.0	2602	6	US-10-453-372-716
23	121	2.0	2617	6	US-10-453-372-666
24	121	2.0	2617	6	US-10-453-372-732
25	121	2.0	2617	6	US-10-453-372-734
26	121	2.0	2617	6	US-10-453-372-736
27	121	2.0	2617	6	US-10-453-372-738
28	121	2.0	2617	6	US-10-453-372-740
29	121	2.0	2617	6	US-10-453-372-742
30	121	2.0	2617	6	US-10-453-372-744
31	121	2.0	2617	6	US-10-453-372-748
32	121	2.0	2617	6	US-10-453-372-750
33	120.5	1.9	2766	6	US-10-877-346-62
34	120	1.9	1717	7	US-11-182-016-20
35	119.5	1.9	216	6	US-10-821-234-984
36	119.5	1.9	6893	7	US-11-205-109-14
37	119	1.9	2617	6	US-10-453-372-746
38	118.5	1.9	1368	7	US-11-043-693-34
39	118	1.9	1709	6	US-10-995-561-973
40	117.5	1.9	919	6	US-10-858-730-206
41	117	1.9	1742	7	US-11-182-016-23
42	116.5	1.9	1439	7	US-11-124-368A-291
43	116.5	1.9	1649	6	US-10-985-561-974
44	114.5	1.9	1560	7	US-11-059-982-1
45	114.5	1.9	2768	6	US-10-510-101-72

ALIGNMENTS

RESULT 1  
US-10-821-234-1209  
; Sequence 1209, Application US/10821234  
; Publication No. US20050255114A1  
; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Andarmani, Susan  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia  
; FILE REFERENCE: 821A  
; CURRENT APPLICATION NUMBER: US/10/821,234  
; PRIOR APPLICATION NUMBER: 2004-04-07  
; PRIOR FILING DATE: 2003-04-07  
; NUMBER OF SEQ ID NOS: 1704  
; SOFTWARE: pt\_seq\_genes Version 1.0  
; SEQ ID NO 1209  
; LENGTH: 1189  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; LOCATION: (1)...(1189)  
; OTHER INFORMATION: Xaa = any amino acid or nothing  
US-10-821-234-1209

Alignment Scores:  
Pred. No.: 3.02e-122 Length: 1189  
Score: 1660.50 Matches: 431  
Percent Similarity: 51.3% Conservative: 184  
Best Local Similarity: 35.9% Mismatches: 445  
Query Match: 26.9% Indels: 139

DB: 6 Gaps: 32

US-10-723-552-3\_copy\_2307\_5741 (1-3435) x US-10-821-234-1209 (1-1189)

1 ATGGGTGCCAGGGCAACACAGTATCCATGGACTACCCGA---AGAACAGTTGACTTG 57  
:::|||||  
64 ValGlyValAspGlyLeuIleSerCysProLeuAlaThrArgProLeuThrCysSerLeu 83  
:::|||||  
58 GGAGTGGGACGGGTAAACCCACTCGTCTGTTGTCTATACCTGAGTGGCCAGCACCCCTCTTA 117  
:::|||||  
84 PheAspAlaIlePheSerHisSerPheLeuIleMetProHisCysProThrProIleLeu 103  
:::|||||  
118 GGTAGAGACTTATTGACCAAGATGGAGCACAATAATTTCTTTTGAA----- 162  
|||||  
104 Gly\*\*\*AspLeuLeuAlaLysValLysAlaSerIleThrPheSerCysProLeuGlnPro 123  
|||||  
163 CAAGGGAACACGAAGTGTCTGCAATAACAAACCTATCACTGTGTGTGACCCCTCCAATTA 222  
:::|||||  
124 GluSerProProAlaProProAlaSerProAlaProAspProPhe----- 138  
:::|||||  
223 GATGACGAATATGACATATCTCTCCCTAGTAAGCCTGATCAAAATATACAAATTTCTGG 282  
:::|||||  
139 ---ProGlnTyProLeuProAlaSerLeuIleAsnPro----- 150  
:::|||||  
283 TTGGAACAGTTTCCCAAGCCTGGGCAGAAACCGCAGGGATGGTTTGGCAAAAGCAAGTT 342  
:::|||||  
151 -----SerValAlaAlaHisHis 156  
:::|||||  
343 CCCCACAAAGTTATTCACTGAAGGCCAGTGCACACACAGTGTCACTGACAGTACCC 402  
|||||  
157 AspProIleArgIleGlnLeuLysAspSerSerLysPheProSerValProGlnTyPro 176  
:::|||||  
403 TTGAGTAAGAAGCTCAAGAAGAAATTCGGCGGCATGTCCAAAGATTAATCAACAGGGC 462  
:::|||||  
177 IleSerLeuThrHisGlnLysGlyLeuGlnProIleValAsnLysLeuCysSerCysSer 196  
:::|||||  
463 ATCTAGTTCCTGCTCAATCTCCCTGGAAATCTCCCTGCTACCGGTTAGAAAGCCTGGG 522  
:::|||||  
197 LeuLeuArgProThrHisSerProTyAsnThrProIleLeuProValLysLys---Ser 215  
:::|||||  
523 ACTAATGACTATCGACCACTGACAGAGTGTGAGAGAGTCAATAAACCGGGTGCAGGATATA 582  
|||||  
216 AspGlySerTyArgLeuValHisAsnLeuGlnAlaIleSerGlnAlaValLeuProIle 235  
:::|||||  
583 CACCCACAGTCCCGAACCTTATAACCTCTGTGTGCTCTCCACCCCAACGAGCTGG 642  
|||||  
236 HisProIleValHisAsnProTyProLeuLeuSerLeuValAlaIleThrAsnThrThrLeu 255  
:::|||||  
643 TATACAGTATTGGACTTAAAGAGTCCCTTCTTCCTGCTGAGATTACACCCCACTAGCCAA 702  
|||||  
256 TyrThrAlaIleAspLeuLysAspAlaPhePheThrIleSerLeuHisProAspSerGln 275  
:::|||||  
703 CCACCTTTTGGCTTCGAATGGAGAGATCCAGGTACGGGAAGACCGGSCAGCTCACCTGG 762  
|||||  
276 AsnLeuPheAlaPheThrTipThrAspProAspThrLeuGlnSerGlnGlnLeuThrTrp 295  
:::|||||  
763 ACCGACTGCCCCAAGGTTCAAGAACTCCCGACCATCTTTGACGAAGCCCTACACAGA 822  
|||||  
296 ThrValLeuProGlnGlyPheArgAspSerProSerLeu---AspGluAlaLeuAlaArg 314  
:::|||||  
823 GACCTGGCCAACTTCAGGATCCAAACACCTCAGGTGACCCCTCCCTCAGTACGTGGATGAC 882  
:::|||||  
315 AsnLeuThrSerLeuAsnLeuSerProSerHis-----LeuGlnTyValAspAsp 331  
:::|||||  
883 CTGCTTCTGGGGGGGACCCCAACAGGACTGCTTTAGAAAGGCACGAAGGCCTACTGCTG 942  
:::|||||  
332 LeuLeuLeuCysSerProSerLeuLysAspSerGlnThrHisThrAlaThrAlaLeuLeu 351  
:::|||||  
943 GAA---TTGCTGACCTAGGCTACAGACCTCTGCTAAGAAGGCCAGATTGTCAGGAGA 999  
:::|||||  
352 AsnPheLeuAlaIleLysGlyTyArgValSerProSerLysAlaGlnLeuSerIleSer 371  
:::|||||  
1000 GAGGTAAACATACTTGGGGGTACAGTTTGGGACGGGCAGCGCATGGCTGACGGGACGCG 1059  
:::|||||

Db 372 MetMetThrTyLeuGlyIleGlnLeuSerProGlyAlaGlnAlaMetThrProAlaArg 391  
:::|||||  
1060 AAGAAAACCTAGTCCAGATACCGGCCCAACACAGCCAAACAAATAGAGAGATTTTGTG 1119  
:::|||||  
392 AlaAlaLeuIleGluAsnLeuProProSerSerLysSerGluIleLeuSerPhePro 411  
:::|||||  
1120 GGGACAGCTGATTTTCAGACTGTGGATCCGGGGTTTGGACCTTAGCAGGCCCACTC 1179  
|||||  
412 GlyLeuAlaGlyPhePheArgIleTrpIleSerAsnPheAlaLeuLeuAlaHisProLeu 431  
:::|||||  
1180 TACCGCTTAACCAAGAAAAAGGGGAATTCCTCGGCTCCTGAGCACACAG-----AAG 1233  
:::|||||  
432 TyrGluValAlaLysGlyProProAsnGluProLeuAsnProSerHisAsnIleLeuPro 451  
:::|||||  
1234 GCATTTGATGCTATCAAAAAGGCCCTGCTGAGCGCAGCTGCTGTGGCCCTCCCTGACGTA 1293  
:::|||||  
452 SerPheHisLysLeuGlnThrAlaLeuValThrAlaSerAlaProSerLeuProAspIle 471  
:::|||||  
1294 ACTAAACCTTTTACCTTTATGTGATGAGGTAAAGGAGTAGCCCGGGGAGTTTAAACC 1353  
:::|||||  
472 SerGlnProPheThrLeuTyThrAlaGluSer\*\*\*GlyIleAlaLeuGlyValLeuGly 491  
:::|||||  
1354 CAAACCCCTAGGA-----CCATGGAGAGAGACCTCTCGCTACCTGTCAAAGAAGCTCGAT 1407  
|||||  
492 GlnGlnLysGlyAsnProProSerPheAlaProValAlaTyLeuSerLysGlnLeuAsp 511  
:::|||||  
1408 CCTGTAGCAGTGTGGCCCATATGCTGAGGCTATCGAGCTATCGAGCTGTGGCCATACTGTC 1467  
|||||  
512 AsnThrValLysGlyTyProAlaCysPheLysAlaLeuGluValValAlaSerLeuAla 531  
:::|||||  
1468 AAGGACGCTGACAAATGACTTTGGGACAGAAATATACTGTAATAGCCCCCATGCAATTG 1527  
:::|||||  
532 LeuGluSerArgLysLeuThrPheSerGlnAsnThrThrValHisSerSerHisAsnLeu 551  
:::|||||  
1528 GAGAACATCGTTCGGCAG-----CCCCCAGACCGATGTGATGACCAAC 1569  
:::|||||  
552 GlnAspLeuLeuSerSerGlnAlaValSerSerLeuProProSerArg----- 567  
:::|||||  
1570 GCCCGATGACCCACTATCAAAAGCTGCTTCTCAGAGAGGGTCACGTTGCTGCCACCA 1629  
:::|||||  
568 IleGlnLeuLeuHis---AlaLeuPheIleLysAsnProLysPheSerLeuThrArgSer 586  
:::|||||  
1630 GCCCTCTCAACCTGCCACTCTTCGCTGAAGAGACTGATGAACCACTGACCTCATGAT 1689  
|||||  
587 AlaSerLeuAsnProAlaSerLeuLeuProValSerSerSerLeuPro---ThrHisSer 605  
:::|||||  
1690 TGCCATCAACTATTATTGAGGAGACTGGGTCCCGAAGGACCTTACAGACATACCCGTG 1749  
:::|||||  
606 SerThrAspIleLeuAspHisValGlnProHisPheProAsnThrSerSerGluProLeu 625  
:::|||||  
1750 ACTGAGAGAGTGTAACTGCTTCACTGACGGAAGCAGCTATGTGTGGGAAGGTAAAGAG 1809  
:::|||||  
626 ThrAsnProAsnAspGlnLeuPheIleAspGlySerSerPheGlnAla-----His 642  
:::|||||  
1810 ATGCTGGGGCGGCTGGTGGACGGACCCGACGATCTGGGCAGCAGCTGCGCGAA 1869  
:::|||||  
643 GlnLeuProProArgLeuLeuAspMetGlnLeuPheProLeuThrLys\*\*\*LeuLysPro 662  
:::|||||  
1870 GGAATCTTCAGCACAAAAGGCTGAGCTCATGGCCCTCAGCAAGCTTTTGGCTGGCCGAA 1929  
|||||  
663 GlySerSerGlnLysAlaGluLeuIleAlaLeuThrArgAlaLeuAsnLeuSerLys 682  
:::|||||  
1930 GGGAAATCCATAAATATTATACGACAGCAGGATGCTTGGCAGTCGACACGTATAT 1989  
:::|||||  
683 GlyLysArgValAsnIleTyThrAspSerLysTyAlaTyHisIleProArgSerHis 702  
:::|||||  
1990 GGGCCATCTATAACAAAGGGGTGCTTACCTACCTCAGCAGGAGGGAATAAAGAACAA 2049  
:::|||||  
703 AlaAlaIleTrpGlnLysArgGlyLeuLeuThrAlaLysGlyThrProIleThrAsnGly 722  
:::|||||  
2050 GAGGAAATTCCTAAGCCCTATTAGAACCGGTACATTTACCAAAAGCGTAGCTATTATAC 2109  
:::|||||

Db 723 HisLeuLeuTyr\*\*\*LeuLeuGlnAlaThrHisLeuProAlaLysAlaGlyValIleHis 742  
Qy 2110 TGCTCTGACATCAGAAAGCTAAAGATCTCATATCCAGAGAAACACGATCGCTACCGG 2169  
Db 743 Cys\*\*GlyHisArgThrGlySerAspGluLeuSerLysGlyAsnArgLysThrAspGlu 762  
Qy 2170 GTTCCCAAGCAGCGCCAGCGGTGTAACTTCTGCTCTAATAAGAAATCCCAAGCC 2229  
Db 763 AlaAlaLysGlnAspSerLeuSerProLeuProAlaProIleLeuLeuValThrProAla 782  
Qy 2230 CCAGAACCCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 2289  
Db 783 ValProProArgTyrProProThr-----GluLysSerSerLeuLeuGln 797  
Qy 2290 TTCTCTGAGACTCCGGAAGGACCTGCTATACCTCAGATGGAAGAAATCTCCGCCAC 2349  
Db 798 GlnGlyAlaSerLeuGlnGlyAspTrpIleIleLysAsnGlnLysProValLeuProGln 817  
Qy 2350 AAGAAGGTTAGAAATGTCCACAGATACATCGTCTAACCCACCTAGGAACTAAACAC 2409  
Db 818 GluGlnSerLysGluIleLeuThrProLeuHisGlnProPheHisIleSerAlaCysPro 837  
Qy 2410 CTGCAGCAGTGTGTCAG-----ACATCCCTTATCATGTCTGAGGCTACACGGA 2460  
Db 838 LeuTyrLeuLeuLeuArgProCysPheSerSerProHisLeuPheIleSerLeu----- 855  
Qy 2461 GTGCTGACTCGGTGTCACAAATGTGTGCTGCCAGCTGGTAAATGCTTAATCCTTCC 2520  
Db 856 -----LysAspIleThrSerAsnCyArgileCysSerValThrSerSerGlnGlyAla 873  
Qy 2521 AGAATGCT-----CCAGGAAGAGACTAAGGGAAGCCACCCAGCGCTCCTCTGG 2571  
Db 874 LeuCysProLeuLeuLeuLeuThrTyrGlnLeuArgGlyThrLeuProGlyGluHisTrp 893  
Qy 2572 GAAGTGCATCTCAGTGAAGTAAAGCGGCTAAATACGGAACAATACCTATTGTTT 2631  
Db 894 GlnValAsnPheThrHisMetProProValLys---LysSerLysTyrLeuLeuThrLeu 912  
Qy 2632 GTACACACCTTTTCCAGATGGGTAGAGCTTATCCTACTAAGAAAGAGACTTCAACCGTG 2691  
Db 913 ValAspThrPheSerGly\*\*ValGluAlaPheProThrProSerGlnLysAlaAlaGlu 932  
Qy 2692 GTGCTAAAAAATCTGGAAGAAATTTTCCAAAGATTTGGAATPACCTAAGGTAATAGG 2751  
Db 933 ValSerGlnIleLeuValThrGluIleIleProArgPheGlyLeuProGlySerIleGln 952  
Qy 2752 TCAGACATGCTCAGCTTTTGTGCCCAGTAAAGTACGAGCTGGCCAGATATTCGGG 2811  
Db 953 SerAspAsnSerProSerPheIleSerGlnIleThrGlnGlnValSerGlnSerLeuGly 972  
Qy 2812 ATTGATTGGAACCTGCATTGTGCATACAGACCCCAAGCTCAGACAGGTAGAGAGATG 2871  
Db 973 IleGlnTrpArgLeuHisIleProCysTrpProGlnThrSerGlyLysValGluArgAla 992  
Qy 2872 AATAGAACCAATTAAGAGACCTTACTAATTTGACCGCGGAGCTGCGCTTAATGATTG 2931  
Db 993 AsnGlyIleLeuLysAlaGlnLeuThrLysLeuThrLysLeuVal---GlnLysProTrp 1011  
Qy 2932 ATAGCTCTCTGCTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCT 2988  
Db 1012 ThrSerLeuLeuProIleAlaLeuGluSerIleArgAlaSerProLysAlaProSerPhe 1031  
Qy 2989 CTGACCCCTTATGAATTAATCTACGCGGGGACCC-----CCCCCATTTG 3030  
Db 1032 LeuSerProPheGluLeuIleTyrGlyArgProPheLeuLeuGlnAsnArgProPro--- 1050  
Qy 3031 GTAGAAATCTCTGTACATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3090  
Db 1051 -----SerAsnSerGlnLeuGlyGluTyrLeuProThrValSerLeuMetSerTyr 1067  
Qy 3091 CTCAGGACCTTGTGCTGTGAGACACAGCAGCTGGAGGCACTCCGGGAGCGCTACTCA 3150  
Db 1068 Leu-----LeuCysGlnGlnAlaAspGlnAlaLeuProLysProHisGlu 1082

Qy 3151 GGAGGAGGAGACTTTCAG-----ATCCACATCGTTTCCAAGTGGAGAT 3195  
Db 1083 GlyValSerAsnProLys\*\*\*ThrCysSerProIleProLys-----Asp 1097  
Qy 3196 TCA-----GTCTACGTTAGACGACCGTCAGAGAAACCTCGAGACTCGGTGGAG 3246  
Db 1098 SerLeuSerArgValThrLeuGlnAsnHisArgGlyLeuAspLeuLeuThrAlaGluLys 1117  
Qy 3247 GGCCCTTATCTCGTACTTTTG-----ACCACACCAACGGCTGTG 3285  
Db 1118 GlyGlyLeuCysIlePheLeuGluGluCysCysPheTyrThrAsnGlnSerGlyLeu 1137  
Qy 3286 -----AAAGTCGAAGGAATCTCCACCTGGATCCAT 3315  
Db 1138 ValGlnAspAlaAlaGlyArgIleAsnGluLysAlaSerGlyArgValGlnTrpLeu--- 1156  
Qy 3316 GCATCCCACTTAAACCGCGGCCACCTCCCGATTCGGGGTGGAAAGCGGAAAG----- 3369  
Db 1157 -----ThrProValIleProAspLeuTrpGluAlaGluAlaGlyGly 1170  
Qy 3370 -----ACTGMAAATCCCTTAAGCTTCGCTCCAT 3399  
Db 1171 SerArgGlyGlnGluIleGluThrIleLeuAlaAsnThrValLysProArgLeuTyr 1189

## RESULT 2

US-10-821-234-1083  
; Sequence 1083, Application US/10821234  
; Publication No. US20050255114A1

; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Andarmani, Susan  
; APPLICANT: Tang, Y. Tom

; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia  
; FILE REFERENCE: 821A  
; CURRENT APPLICATION NUMBER: US/10/821,234  
; CURRENT FILING DATE: 2004-04-07  
; PRIOR APPLICATION NUMBER: US 60/462,047  
; PRIOR FILING DATE: 2003-04-07  
; NUMBER OF SEQ ID NOS: 1704  
; SOFTWARE: pt\_seq\_genes Version 1.0  
; SEQ ID NO 1083  
; LENGTH: 1150  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1150)  
; OTHER INFORMATION: Xaa = any amino acid or nothing

US-10-821-234-1083

Alignment Scores:  
Pred. No.: 1,86e-118 Length: 1150  
Score: 1612.00 Matches: 421  
Percent Similarity: 52.3% Conservative: 178  
Best Local Similarity: 36.8% Mismatches: 444  
Query Match: 26.1% Indels: 102  
DB: Gaps: 35

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-10-821-234-1083 (1-1150)

Qy 82 TTCTCTGTCATCATCGTGTGCTCCAGCACCCCTCTTAGGTAGAGACTTTATTGACCAAGATG 141  
Db 59 PheLeuCysProGluCysProThrProLeuLeuGlyArgAspLeuLeuSerLysMet 78  
Qy 142 GGAGCAAAATTTCTTTTGAACAAGGAACCAAGAGTGTCTCAAAAT---AACAAACCT 198  
Db 79 GlyAlaThrIleSerLeuGluGluAspArgLeuGlnValGluAlaGluProGluGlnGly 98  
Qy 199 ATCACTGTGTGTACCTCCCAATTAGATGACGAATATCGACTATCTCTCCCTAGTAAG 258  
Db 99 IleHisLeuLeuAlaLeuAsnGlyGlnGluLeuGlu----- 111



```
Db 810 -----*TrpArgAspValLeuTrpAsnSerValHisLeuLysPheThrArg 825
Qy 2455 CCAGGAGTGGTGAATCGGTGGTCAAAATGTTGTCCTCCAGCTGGTAAATGCTAAT 2514
Db 826 LysLeuLeuArgLysSerValLysGlnValThrPheValSerGluLeuCysSerArgAsn 845
Qy 2515 CTTCCAGA-----ATGCCTCCAGGGAAG-----AGACTAAGGGGAAGC 2553
Db 846 AsnProHisThrHisProIleProProSer***LeuLysLeuValGlnHisArgGlyThr 865
Qy 2554 CACCCAGCGCTCACTGGGAAGTCACTTCACTGAGGTAAAGCCGGCTAAATACGGAAC 2613
Db 866 TyrLeuGlyGluAspTrpGlnValValLeuThrGlnMetThrPro---AsnLeuGlyTyr 884
Qy 2614 AAATACCTATTGGTTTTGTAGACACCTTTTTCAGATGGGTAGAGGCTTATCCTACTAAG 2673
Db 885 LysTyrLeuLeuValPheValAspThrPheThrGlyArgValLysGlyPheProThrCys 904
Qy 2674 AAAGAGACTTCAACCGTGGTGGCTTAAATAAATACTCGAAGAAATTTTTCAGAGTTTGA 2733
Db 905 ThrGlyLysAlaValGluValCysLysProSerLeuLysGluValIleSerGlnPheGly 924
Qy 2734 ATACCTAAGGTAATAGGTGACAAATGCTCCAGCTTTTGTGCCAGGTAAAGTCAGGGA 2793
Db 925 LeuProLysSerProGlnSerGlyAsnArgLeuSerPheMetGlyLysIleThrGlnSer 944
Qy 2794 CTGGCCAAAGATATTTGGGATTTGGAACTGCTATGTCATACAGACCCCAAGCTCA 2853
Db 945 LeuSerThrThrLeuGlyIleAspTyrGluLeu-----AlaProGlnSerSer 960
Qy 2854 GGACAGGTAGAGAGTGAATAGAACCACTTAAAGAGACCCCTTACTAAATTCACCGCGAG 2913
Db 961 GlyLysVal---LysMetAsnHisThrLeuGluThrThrLeuAlaLysLeuPheGlnGlu 979
Qy 2914 ACTGGCGTTAATGATGATAGTCTCTGTCGCCCTTGTGCTTTTGTAGGGTTAGGAACACC 2973
Db 980 MetHis---GluSerTrpValLysMetLeu---LeuSerLeuLeuArgValArgAla 997
Qy 2974 CTTGGA---CAGTTTGGCTGACCCCTATGAAATTTACTCTACGGGGACCCGCCCATG 3030
Db 998 ProSerCysSerLeuArgLeuSerSerGlnMetIleTyrGlnArgPro----- 1014
Qy 3031 GTAGAAATGCTTCTGTACATAGTGTGCTGCTGCTTTTCCAGCCCTTTGTTCTPAGG 3090
Db 1015 -----PheLeuThrThrAspLeuLeuLeuGlyGluGluLeuHisLysLys 1029
Qy 3091 CTCAGGCACTTGAGTGGTGAGACACAGAGCGGTGGAGGCACTCCGGGAGGCCCTACTCA 3150
Db 1030 Leu-----GlnTyrArgPheArgThrAsp 1037
Qy 3151 GGAGGAGGAGACTTGCAGATC-----CCACATCGT----- 3180
Db 1038 SerLysGlyAspPheArgIleTyrLysGlnAsnProAlaSerProHis**GlyTrpGlu 1057
Qy 3181 -----TTCCAAGTG-----GGAGATTCACTTACGTGTAGAGCCACCGTCGAGGAAC 3228
Db 1058 AsnProPheGlnValAsnProGlyAspGlnValLeuLeuLysAla**ArgAlaGlySer 1077
Qy 3229 CTCGAG-----ACTCGGTGGAAGGCCCTTATCTCGTACTTTTTCACACACCA 3276
Db 1078 ProGluAspHisProLeuLeuLysTrpGluGlyProCysTrpValIleLeuThrThrPro 1097
Qy 3277 ACGGCTGTGAAGTCGAGGAATCTCCACTGGATCCATGATCCATCCAGCTTAACCGCG 3336
Db 1098 ThrAlaAlaAsnSerGlnGlyIleThrSerTrpValHisLeuSerArgSerGluMetLeu 1117
Qy 3337 CCACCT-----CCCGATTTCGGGGTGGGAAGCCGAAAGACTGAAAT 3378
Db 1118 SerProLysCysPheGlnThrArgProAspGlyProPheTyrSer**LysProValGlu 1137
Qy 3379 CCCTTAAAGCTTCGC 3393
|||||
|||||
```

```
Db 1138 AspLeuLysPheArg 1142
RESULT 3
US-11-042-988-12
; Sequence 12, Application US/11042988
; Publication No. US20050244818A1
; GENERAL INFORMATION:
; APPLICANT: SILICIANO, ROBERT
; APPLICANT: ZHANG, HAILI
; APPLICANT: ZHOU, YAN
; TITLE OF INVENTION: SINGLE CELL ANALYSIS OF HIV REPLICATION CAPACITY AND
; TITLE OF INVENTION: DRUG RESISTANCE
; FILE REFERENCE: 62760(71699)
; CURRENT FILING DATE: 2005-01-25
; PRIOR APPLICATION NUMBER: 60/540,716
; PRIOR FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 12
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-11-042-988-12
Alignment Scores:
Pred. No.: 3,9e-26 Length: 912
Score: 431.00 Matches: 209
Percent Similarity: 35.9% Conservative: 145
Best Local Similarity: 21.2% Mismatches: 345
Query Match: 7.0% Indels: 286
DB: 39 Gaps: 39
US-10-723-552-3_COPY_2307_5741 (1-3435) x US-11-042-988-12 (1-912)
Qy 85 CTGGTCATACCTGAGTGCACCCAGCACCCCTCTTAGGTAGAGACTTATTGACCAAGATCGGA 144
Db 40 ValLeuValGlyProThrProValAsnIleGlyArgAsnLeuLeuThrGlnIleGly 59
Qy 145 GCACAAATTTCTTTTGAACAAGGGAAACAGAGTGTCTGCAAAATAACAACCTATCACT 204
Db 60 CysThrLeuAsnPhe-----ProIle--- 66
Qy 205 GTCTTGACCCCTCCAATTAGATGACGAATATCGACTATCTCCCTTAGTAAAGCCTGAT 264
Db 67 -----SerProIle----- 69
Qy 265 CAAAATATACAAATTCGTGTTGGAACAGATTTCCTCCAGCCCTGGGCGAGAAACCGCAGGATG 324
Db 69 ----- 69
Qy 325 GGTTCGCAAGCAAGTTCCTCCACCAAGTTATTCAACTGAAGCCAGTGCACACCAAGTG 384
Db 70 -----GluThrValPro-----ValLysLeuLysProGlyMetAspGlyPro 83
Qy 385 TCAGTCAGACAGTACCCCTTGTAGTAAAGAGCTCAAGAGGAATTCGGCGCATGTCCAA 444
Db 84 LysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThr 103
Qy 445 AGATTAAATCCACAGGGCATCTTAGTTCCTGTC-----CAATCTCCTCGAATATCTCCC 498
Db 104 GluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrPro 123
Qy 499 CTGCTACCGGTTAGAAAGCTGGGACTAATGACTATCGACCATACAGGACTTGAGAGAG 558
Db 124 ValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGlu 143
Qy 559 GTCATAAAGCGGTGAGGAT-----ATACACCAACAGCTCCGCAACCTTATAAC 609
Db 144 LeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGly 163
Qy 610 CTCTTGCTGTCTCTCCACCCCAACCGAGCTGGTATATACAGTATTGAGCTTAAAGGATGCC 669
|||||
|||||
```



[illegible]

1663	QY	---	GAGACTGATGAAC	CACAGTGCATCATGATATGGCATCAACATATATGATTGAGGAGACTGGG	1719	
480	Db	PheVal	AsnThrProProLeuValLysLeuLeuTyrTyrGlnLeuGluLysGlu	---	496	
1720	QY	GTCCGCAAGGACCTTACAGACATACCGCTCAGCTGGGAAGTGTCTAACCTCGTTCACTGCATGC	1779			
497	Db	---	ProIleValGly	---	AlaGluThrPheTyrValAsp	507
1780	QY	GGAAAGCAGC	---	TATGTGTGGGAAGGTAAAGAG	1809	
508	Db	GlyAlaAlaAsnArgGluThrLysLeuGluLysAlaGlyTyrValThrAsnArgGlyArg	527			
1810	QY	ATGCTGGGGCGGGTGGTGGACCGGACCGCACAGATCTGGGCCAGCACCTGCCGGAA	1869			
528	Db	GlnLysValValThrLeuThrAsp	---	---	---	535
1870	QY	GGAACTTCAGCACAAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGGCGCTGGCGCAA	1929			
536	Db	---	ThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGln	---	AspSer	553
1930	QY	GGGAAATCCATAACATTTATACGGACAGCAGGATATGCTTCTGGCATGTCACACGCTACAT	1989			
554	Db	GlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGly	---	---	---	568
1990	QY	GGGGCCATCTATAACAAGGGGTGCTTACCTCAGCAGGGAGGGAATAAAGAACAAA	2049			
569	Db	---	IleIleGlnAlaGlnProAspGlnSerGluSer	579		
2050	QY	GAGGAAATTCCTAAGCCTATTAGAACCGGTATCATTTACCAAAAAGGCTAGCTATTATACAC	2109			
580	Db	GluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTyr	599			
2110	QY	TGTCTCGCATCAGAAAGCTAAAGATCTCATATCCAGAGAAACCATGCTGCTACCCG	2169			
600	Db	ValProAlaHisLysGlyIle	---	GlyGlyAsnGluGlnValAspLys	614	
2170	QY	GTTCCCAAGCAGGCGCCAGGGGTGTAACTCTCTGCCTATAAGAAATGCCCAAGCC	2229			
615	Db	LeuValSerAlaGlyIleArgLysValLeuPheLeu	---	---	---	626
2230	QY	CCAGAACCCAGACAGCATGACCCCTAGAAGACTGGCAGAGATATAAAGATAGACCAG	2289			
626	Db	---	---	---	---	626
2290	QY	TTCTCTGAGACTCCGGAAAGGACCTGTCTATACCTCAGATGGGAAGGAAATCCTGCCCCAC	2349			
626	Db	---	---	---	---	626
2350	QY	AAAGAAGGGTTAGATATATGTCACAGACATACATGCTCTAACCCACCTAGGAACCTAAACAC	2409			
627	Db	---	AspGlyIleAspLysAlaGlnAspGluHisGluLysTyrHisSerAsnTyrArgAla	645		
2410	QY	CTGACAGAGTTGTCAGAACATCCCTTATCATGTTCTGAGGCTACACAGAGTGGCTGAC	2469			
646	Db	Met	---	ProProValValAlaLys	658	
2470	QY	TCGGTGTCAAAACATTTGTGTCCTGCCCTGCTGTTAATGCTAATCCTTCCAGAATGCCT	2529			
659	Db	GluIleValAlaSerCysAspLysCysGlnLeu	---	---	---	669
2530	QY	CCAGGGAAGAGACTAAGGGGAAGC	---	CACCAGCGGCTCACTGGGAGTGGAC	2580	
670	Db	LysGlyGluAlaMetHisGlyGlnValAspCysSerProGlyIle	---	TrpGlnLeuAsp	688	
2581	QY	TTCACTGAGGTAAAGCCGGCTAAATACGGAACAAATACCTATTGTTGTTTGTAGACACC	2640			
689	Db	CysThrHisLeuGlu	---	GlyLysValIleLeuValAlaHisVal	703	
2641	QY	TTTTCCAGATGGGTAGAGGCTTATCTACTAAGAAAGAGACTTCAACCCGTGGTGGCTAAA	2700			
704	Db	AlaSerGlyTyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyr	723			



```
QY 1717 GGGGTCCGCAAGGACCTTTACAGACATACCGCTGACTGGAGAAGTGCTAACTGGTTCACT 1776
Db 435 -----ProlineValGly-----AlaGluThrPheTyrVal 444
QY 1777 GACGGAAGCAGC-----TATGTGGTGGAAAGGTAAG 1806
Db 445 AspGlyAlaAenArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGly 464
QY 1807 AGGATGGCTGGGCGCGGTGGTGGACGGGACCGCCACGATCTGGGCCAGCAGCCTGCCG 1866
Db 465 ArgGlnLysValValThrLeuThrAsp----- 473
QY 1867 GAAGGAACCTTCAGACAAAGCGCTAGCTCATGGCCCTCACGAAGCTTTTCGGCTGSGCC 1926
Db 474 -----ThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGln---Asp 490
QY 1927 GAAGGGAATCCATAAATATATACGACGACGAGGTATGCTTTGGCAGCTGCACAGTA 1986
Db 491 SerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGly----- 506
QY 1987 CATGGGCGCATCTATAAACAAGGGGGTGTCTTACCTCAGCAGGAGGGAATAAAGAAC 2046
Db 507 -----IleIleGlnAlaGlnProAspGlnSerGlu 516
QY 2047 AAAGAGGAATCTAAGCCTATTAGAAGCCGTACATTTTACCAAAAGGCTAGCTATTATA 2106
Db 517 SerGluLeuValAsnGlnIleIleGluLeuIleLysLysGluLysValTyrLeuAla 536
QY 2107 CACTGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACAGATGGCTGAC 2166
Db 537 TrpValProAlaHisLysGlyLe-----GlyGlyAsnGluGlnValAsp 551
QY 2167 CGGGTTCGAAGCAGGAGCCAGGGGTGT 2196
Db 552 LysLeuValSerAlaGlyIleArgLysVal 561

RESULT 5
US-10-507-928-12
; Sequence 12, Application US/10507928
; Publication No. US2005026024A1
; GENERAL INFORMATION:
; APPLICANT: POWDERMED LIMITED AND GLAXO GROUP LIMITED
; TITLE OF INVENTION: ADJUVANT
; FILE REFERENCE: N.88232B GCW
; CURRENT APPLICATION NUMBER: US/10/507,928
; CURRENT FILING DATE: 2004-09-17
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of the coding insert in p73i-RT3
US-10-507-928-12

Alignment Scores:
Pred. No.: 5,07e-16 Length: 562
Score: 301.00 Matches: 145
Percent Similarity: 37.2% Conservative: 107
Best Local Similarity: 21.6% Mismatches: 264
Query Match: 4.9% Indels: 154
Db: 6 Gaps: 26

US-10-723-552-3_COPY_2307_5741 (1-3435) x US-10-507-928-12 (1-562)
QY 322 ATGGGTTTGGCAACCAAGTATCCCAACAGTATTTCACCTGAAGCCAGCGCCACACCA 381
Db 1 MetGlyProIleSerProIleThrValSerValLysLeuLysLeuPheProGlyMetAspGly 20
QY 382 GTGTCAGTCACAGCATCCCTTTAGTAAAGAGCTCAAGAGGAATTCGGCGCATGTC 441
Db 21 ProlLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCys 40
```

```
QY 442 CAAAGATTAATCCAAACAGGCGATCCTAGTTCTCTGTC-----CAATCTCCCTGGAATACT 495
Db 41 ThrGluMetGluLysGluGlyLysIleSerLysLysIleGlyProGluAsnProTyrAsnThr 60
QY 496 CCCCTGCTACCGGTTAGAAAAGCCCTGGGACTAATGACTATCGACGACGTACAGCACTTGAGA 555
Db 61 ProValPheAlaIleLysLysLysAspSerThrLysIleArgLysLeuValAspPheArg 80
QY 556 GAGTCTAATAAACCAGGTCGAGGAT-----ATACACCCCAACAGTCCCGAACCCCTTAT 606
Db 81 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 100
QY 607 AACCTCTTGTGTCTCTCCACCCCAACGAGGAGCTGTATACAGTATTGGACTTAAAGGAT 666
Db 101 GlyLeu-----LysLysLysLysSerValThrValLeuAspValGlyAsp 115
QY 667 GCCTTCTTCTGCTGAGATTACACCCCACTAGCCCAACCCACTTTTTCCTTCGAATGGAGA 726
Db 116 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 135
QY 727 GATCCAGGTACGGGAAGAACCGGG---CAGCTCACCTGGACCCGACCTGCCCCAGGGTTC 783
Db 136 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 155
QY 784 AAGAACTCCCGCCGACCATCTTTTGCAAGCCCTACACAGAGACCTGGCCAACTTCAGGATC 843
Db 156 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 175
QY 844 CAACACCTCTCAGGTGACCTCTCTCAGTACGTGATGACCTCTTCTGGCGGAGCCACC 903
Db 176 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrVal---GlySerAsp 194
QY 904 AAACAGGACTCTTAGAAGGCACGAAG---GCATCTACTGCTGGAATTTGCTGCACCTA 957
Db 195 LeuGluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrp 214
QY 958 GGCTACAGAGCCTCTGCTAAGAGCCAGATTTCAGGAGAGAGGTAACACTTGTG--- 1014
Db 215 GlyLeuThrThrProAspLysLysHisGln-----LysGluProProPheLeuTrp 231
QY 1015 ---GGGTACAGTTTTCGGGACGGGACGCGTGGCTGACGGAGGACGCGAAGAAACTGTA 1071
Db 232 MetGlyTyrGluLeuHisPro---AspLysTrp-----ThrVal 243
QY 1072 GTCCAGATACCGGCCCCCAACC-----ACGCCAAACAAATAGAGAGTTTGTG 1119
Db 244 GlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuVal 263
QY 1120 GGGACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTT----- 1158
Db 264 GlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLys 283
QY 1159 -----GCGACCTTAGCAGCCCACTCTACCCGCTAACCAAA----- 1194
Db 284 LeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeu 303
QY 1195 -----GAAAAGGGGAAATTCCTGGGCTCTCTGAGCACCAGAGGCATTTGATGCT 1245
Db 304 GluLeuAlaGluAsnArgGluIleLysGluProValHisGlyValTyrTyrAspPro 323
QY 1246 ATCAAAAAGGCCCTCTGTGAGCGCACCTGCTCTGGCCCTCCCTGACGTAACCTAAACCCCTT 1305
Db 324 SerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTrpThr-----Tyr 341
QY 1306 ACCCTTTATGTGGATAGCGTAAGGAGTAGCCCGGGAGGATTTTAACCCAAACCCCTAGGA 1365
Db 342 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 361
QY 1366 CATGGAGAGACCTGTGCGCTACTGTCAAGAGCTCGATCTCTGTAGCAGCAGTGGTGG 1425
Db 362 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr----- 379
```

```
QY 1426 CCCATATCCCTGAAGGCTATCGCAGCTGTGGCCATCTGTCAGGAGCGCTGACAAATTG 1485
Db 380 -----GluSerIleValIleTrpGlyLys----- 387
QY 1486 ACTTTGGACAGATATAACTGTATAGCCCCCATGCAATGGAGAACATGTTGGCGAG 1545
Db 388 -----ThrProLysPheLysLeuProIleGlnLysGlu 398
QY 1546 CCCCAGACCGATGATGACCAACGCCGCGATGACCCACTATCAAGCTCTTCTCACA 1605
Db 399 ThrTrpGluThrTrpTrpThrGlu-----TyrTrpGln----- 409
QY 1606 GAGAGGTCAGCTTCGCTCCACAGCCGCTCTCAACCCCTGCCACTCTTCTGCTGAA--- 1662
Db 410 -----AlaThrTrpIleProGluTrp 416
QY 1663 -----GAGATGATGAACAGTACTCATGATTCGCAATCAATATTGATTGAGGAGACT 1716
Db 417 GluPheValAsnThrProProLeuValLysLeuTrpIleGlnLeuLysGlu----- 434
QY 1717 GGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGGAGAGTGTCTAAACCTGGTTCACT 1776
Db 435 -----ProIleValGly---AlaGluThrPheTyrVal 444
QY 1777 CACGGAAGCAGC-----TATGTGTGGAAGGTAAAG 1806
Db 445 AspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGly 464
QY 1807 AGGATGGTGGGGGGGTGGAGCGGGAGCCCGCAGCATCTGGGCCAGCAGCTGCCG 1866
Db 465 ArgGlnLysValValThrLeuThrAsp----- 473
QY 1867 GAAGGAACCTCAGCACAAAGGCTGAGCTCATGCGCCCTCAGCAAGCTTTTCGGCTGGCC 1926
Db 474 -----ThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGln---Asp 490
QY 1927 GAAGGAAATCCATAACATTTATACGACAGAGGTATGCTTTGGGACTGGCACGTA 1986
Db 491 SerGluLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGly----- 506
QY 1987 CATGGGCCATCTATAACAAAGGGGTGTCTTACCTCAGCAGGAGGGAATAAAGAAC 2046
Db 507 -----IleIleGlnAlaGlnProAspGlnSerGlu 516
QY 2047 AAAGAGGAATTTAAGCTATTAGAAGCCGTACATTTTACCAAAAGGCTAGCTATTATA 2106
Db 517 SerGluLeuValAsnGlnIleIleGluLeuLysLysGlyLysValTyrLeuAla 536
QY 2107 CACTGTCTTGACATCAGAAAGCTTAAGATCTCATATCCAGAGAAACCGATGGCTGAC 2166
Db 537 TrpValProAlaHisLysGlyIle-----GlyGlyAsnGluGlnValAsp 551
QY 2167 CGGTTGCCAGCAGGCGCCAGCGGTGTT 2196
Db 552 LysLeuValSerAlaGlyIleArgLysVal 561
```

## RESULT 6

```
US-11-029-465-10
; Sequence 10, Application US/11029465
; Publication No. US20050256070A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Ralph P.
; APPLICANT: Thomsen, Lindy
; APPLICANT: Van-Wely, Catherine
; APPLICANT: Ertl, Peter
; TITLE OF INVENTION: Adjuvant
; FILE REFERENCE: 033267-015
; CURRENT FILING DATE: 2005-01-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 562
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of RT insert of p7077-RT3
US-11-029-465-10

Alignment Scores:
Pred. NO.: 5,07e-16 Length: 562
Score: 301.00 Matches: 145
Percent Similarity: 37.2% Conservative: 104
Best Local Similarity: 21.6% Mismatches: 267
Query Match: 4.9% Indels: 154
DB: 7 Gaps: 26

US-10-723-552-3_COPY_2307_5741 (1-3435) x US-11-029-465-10 (1-562)
QY 322 ATGGGTTTGGCAAGCAAGTTCCTCCACAGTTATTCACTGAGGCGCAGTGCCACACCA 381
Db 1 MetGlyProIleSerProIleGluThrValSerValLysLeuLysProGlyMetAspGly 20
QY 382 GTGTGCTCAGACAGTACCCCTTGCAGTAAAGAGCTCAAGAGGAATTCGGCCGCATGTC 441
Db 21 ProLysValLysGlnTrpProLeuThrGluLysLysLysAlaLeuValGluLysCys 40
QY 442 CAAAGATTAATCCACAGGCGCATCTAGTTCTCTGTC-----CAATCTCCCTGGAATACT 495
Db 41 ThrGluMetGluLysGluGlyLysIleSerLysLysGlyProGluAsnProTyrAsnThr 60
QY 496 CCCCTGCTACCGGTAGAAAGCCCTGGAGCTAATCACTATCGACAGTACAGACTTGAGA 555
Db 61 ProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 80
QY 556 GAGGTCAATAAACCAGGTCAGAGAT-----ATACACCCCAACAGTCCCAACCCCTTAT 606
Db 81 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 100
QY 607 AACCTCTTGTGTCTCTCCACCCCAACGAGAGTGTATACAGTATTGGACTTAAAGAT 666
Db 101 GlyLeu-----LysLysLysSerValThrValLeuAspValGlyAsp 115
QY 667 GCCTTCTTCTGCTGAGATTACACCCCTAGCCCAACCACTTTTGTGCTTCGAATGGAGA 726
Db 116 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 135
QY 727 GATCCAGGTACGGGAAGAACCGGG---CAGCTCACCTGGAGCCCGCTGCCCCAAGGGTTC 783
Db 136 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 155
QY 784 AAGAACTCCCGACCATCTTTTGACGAAGCCCTACACAGAGACCTGGCCAACTTCAGGATC 843
Db 156 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 175
QY 844 CAACACCTCAGGTGACCCCTCCAGTACGTGGATGACCTGCTTCTGGGGGAGCCACC 903
Db 176 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrVal---GlySerAsp 194
QY 904 AAACAGGACTGCTTAGAAGGCACGAAG-----GCACCTACTGCTGGAATGTCTGACCTA 957
Db 195 LeuGluIleGlyGlnHisArgThrLysIleGluLeuLeuArgGlnHisLeuLeuArgTrp 214
QY 958 GGCTACAGAGCCCTCTGCTAAGAAAGCCCGCAGATTTGCAGGAGAGAGAGGTAAACATCTTG--- 1014
Db 215 GlyLeuThrThrProAspLysLysHisGln-----LysGluProProPheLeuTrp 231
QY 1015 ---GGGTACAGTTTGGGGACCGGCGACCGATGCTGACGGAGGACCGAAGAAACCTGTA 1071
Db 232 MetGlyTyrGluLeuHisPro---AspLysTrp-----ThrVal 243
QY 1072 GTCAGATACCGGCCCAACC-----ACAGCCAAACAAATGAGAGAGTCTTTTG 1119
Db 244 GlnProIleValLeuProGluLysAspSerThrValAsnAspIleGlnLysLeuVal 263
QY 1120 GGGACAGCTGGATTTTGCAGACTGTGATCCCGGGTGT----- 1158
```



Db	155	LysGlySerProAlaIlePheGlnSerMetThrLysIleLeuGluProPheArgLys	174
Qy	844	CAACACCTCAGGTGACCTCTCTCCAGTACGTGATGATGAGGTAAACATACTTGGGGGTACAGT	1023
Db	175	GlnAsnProAspIleValIleTyrGlnIleTyrMetAspLeuTyrValGly	191
Qy	904	AAACAGGACTCTTAGAAGGCACAAAGGCACACTACTGTGTGAATTTGTCTGACCTAGGTCTAC	963
Db	192	-----SerAspLeuGluIle	196
Qy	964	AGAGCTCTGTAAGAAGCCACAGATTGTCAGGAGAGGTAAACATACTTGGGGGTACAGT	1023
Db	197	GlyGlnHisThrArgLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThr	216
Qy	1024	TTGGGGACGGCGCACGA-----TGCGTG-----ACGGAG	1053
Db	217	ThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHis	236
Qy	1054	GCACGGAAGAAACTGTAGTCCAGATACCGGCCCAACC-----ACAGCCAAA	1101
Db	237	ProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsn	256
Qy	1102	CAATGAGAGAGTTTTGGGACACAGCTGGATTTTGCAGACTGTGGATCCCGGGGTTT---	1158
Db	257	AspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLys	276
Qy	1159	-----CGGACCTTAGCAGCCCACTCTACCCGCTA	1188
Db	277	ValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeu	296
Qy	1189	ACCAAA-----GAAAGGGGAATTTCTCTGGGTCTCTGAGCAC	1227
Db	297	ThrGluGluAlaGluLeuLeuAlaGluAsnArgGluIleLeuLysGluProValHis	316
Qy	1228	CAGAAGCATTTGATGTATCAAAAGGCCCTGCTGAGCGCACCTGCTGTGGCCCTCCCT	1287
Db	317	GlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGly	336
Qy	1288	GACGTAACTAAACCTTTTACCCTTTTATGTGATGAGGTAGGAGTAGCCCGGGAGTT	1347
Db	337	GlnTrpThr-----TyrGlnIleTyrGlnGluProPheLysLeuLeuLysThrGlyLys	354
Qy	1348	TTAACCCAAACCTTAGGACCATGAGAAGACCTCTGCTGCTACCTGTCAAAAGAGCTCGAT	1407
Db	355	TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln	374
Qy	1408	CTGTAGCAGTGGTGGCCCATGATGCTGAGGCTATCGCAGCTGTGGCCATCTAGTGGT	1467
Db	375	LysIleThrThr-----GluSerIleValIleTrpGly	385
Qy	1468	AAGGACCTGCATAATTTGACTTTGGGACAGATATACTGTATAGCCCCCATGCATTG	1527
Db	386	Lys-----ThrProLysPheLys	391
Qy	1528	GAGAACATCTGTTGGGCAGCCCCCAGACCGATGGATGACCAAGCCCGCATGACCCACTAT	1587
Db	392	LeuProIleGlnLysGluThrTrpGluThrTrpTrpGlu-----TyrTrp	407
Qy	1588	CAAGCCTGCTTCTCAGAGAGGGTCACTGCTCCACAGCCGCTCTCAACCTGCTGCC	1647
Db	408	Gln-----Ala	409
Qy	1648	ACTCTTCTGCTGAA-----GAGACTGATGAACACGACTGACTCATGATTGCCATCAA	1698
Db	410	ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln	429
Qy	1699	CTATTGATTGAGGAGACTGGGGTCCGCAAGGACCTTACAGACATACCGCTGACTGAGAA	1758
Db	430	LeuGluLysGlu-----ProIleValGly---	437
Qy	1759	GTGCTAACCTGGTTTCACTGACGGAAGCAGC-----	1788

Db 438 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly 457  
 QY 1789 TATGTGTGGGAAGGTAAGAGCATGGCTGGGGCGCGGTGGTGGACGGGACCCGACCATC 1848  
 Db 458 TyrValThrAsnArgGlyArgGlnLysValThrLeuThrAsp----- 472  
 QY 1849 TGGCCACGACGCTCCCGAAGGAGGATTCACGACAAAGGCTGAGCTCATGGCCCTCACG 1908  
 Db 473 -----ThrThrAsnGlnLysThrGluLeuGlnAlaIleTyr 484  
 QY 1909 CAAGCTTTGGCGCTGGCCGAAGGAATTCATAACATTTATACGACACAGCATGATGCC 1968  
 Db 485 LeuAlaLeuGln---AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAla 503  
 QY 1969 TTTGCGACTGCACACGATACATGGGGCCATCTATATAACAAGGGGTGCTTACCTCAGCA 2028  
 Db 504 LeuGly-----IleIleGlnAla 509  
 QY 2029 GGGAGGGAAATAAAGAAACAAGAGGAAATCTAAGCCTATTAGAAGCGGTACATTACCA 2088  
 Db 510 GlnProAspGlnSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLys 529  
 QY 2089 AAAAGGCTAGCTATTATACACTGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGA 2148  
 Db 530 GlnLysValTyrLeuAlaTyrValProAlaIleLysGlyIle-----Gly 544  
 QY 2149 GGAAACAGATGGCTGACCGGTTGCCAAGCAGGCGACCCGAGGTGT 2196  
 Db 545 GlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 560

RESULT 8  
 US-11-022-562-221  
 ; Sequence 221, Application US/11022562  
 ; Publication No. US20050249742A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruprecht, Ruth M.  
 ; APPLICANT: Shisong, Jiang  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING  
 ; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE  
 ; FILE REFERENCE: DPN-043CN  
 ; CURRENT APPLICATION NUMBER: US/11/022,562  
 ; CURRENT FILING DATE: 2004-12-22  
 ; PRIOR APPLICATION NUMBER: PCT/US03/20322  
 ; PRIOR FILING DATE: 2003-06-27  
 ; PRIOR APPLICATION NUMBER: 60/392718  
 ; PRIOR FILING DATE: 2002-06-27  
 ; NUMBER OF SEQ ID NOS: 340  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 221  
 ; LENGTH: 210  
 ; TYPE: PRT  
 ; ORGANISM: Human Immunodeficiency Virus  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: 31, 97, 140, 141, 144, 178  
 ; OTHER INFORMATION: Xaa = Any Amino Acid  
 US-11-022-562-221

Alignment Scores:	
Pred. No.:	1.35e-09
Score:	217.50
Percent Similarity:	47.9%
Best Local Similarity:	30.5%
Query Match:	3.5%
DB:	9

US-10-723-552-3\_COPY\_2307\_5741 (1-3435) x US-11-022-562-221 (1-210)  
 QY 478 CAATCTCCCTGGAACTACTCCCTGCTACCGGTAGAAAGCCTGGGACTAATGACTATCGA 537  
 Db 16 GlnAsnProTyrAsnThrProValPheAlaIleLysLysLysAsp\*\*\*ThrLysTrpArg 35  
 QY 538 CCAGTACAGACTTTGAGAGAGGTCAATAAAGGGGTGCAGAT-----ATACACCCA 588

```
Db 36 LysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeu 55
    : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 589 ACAGTCCCGAACCCCTTATAACCTCTTGCTGCTCTCCACCCACCGAGAGCTGTATACA 648
    : : : : : | | | | | : : : : : | | | | | : : : : :
Db 56 GlyIleProHisProAlaGlyLeu-----LysLysLysSerValThr 70
    : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 649 GTATTGGACTTAAAGATGCTCTTCTGCTGCTGAGATTACACCCACCTAGCAACCACTT 708
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 71 ValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyr 90
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 709 TTTGCTTTTGAATGAGAGATCCAGGTACGGGAAGAACCGGG---CAGCTCACCTGGAC 765
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 91 ThrAlaPheThrIlePro**ThrAsnAsnGluThrProGlyIleArgTyrGlnTyrAsn 110
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 766 CGATGCCCCAAGGGTTCAAGAACTCCCGGACCATTCTTTGAGGAAGCCCTACACAGAC 825
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 111 ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIle 130
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 826 CTGGCCAACTTCAGATCCAAACCTCAGGTGACCTCTCCAGTACGTGATGACCTG 885
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 131 LeuGluProPheArgIleLysAsnPro*****ValIle**GlnTyrMetAspAspLeu 150
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 886 -----CTTCTGCGGGAGCCACCAACAGGACTGCTTAGAAGGCACCGAAG 930
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 151 TyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluLeuArgLys 170
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 931 GCATCTACTGTGGAATTTCTGACTAGCTACAGAGCCTCTGTAAAGAGCCGAGATT 990
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 171 HisLeuLeu-----SerTrpGlyPhe**ThrProAspLysLysHisGln--- 185
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 991 TGCAGGAGAGGTAACATCTTG-----GGTACAGTTTGGCGGACGGCGCGATGG 1044
    : : : : : | | | | | : : : : : | | | | | : : : : :
Db 186 -----LysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro---AspLysTrp 202
    : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 1045 CTGACGGAGCGGACCGAAGAACTGTAGTCCAGATACCG 1083
    : : : : : | | | | | : : : : : | | | | | : : : : :
Db 203 ThrValGlnPro-----IleGlnLeuPro 210
```

## RESULT 9

```
US-11-230-251-26
; Sequence 26, Application US/11230251
; Publication No. US20060019322A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Hervé
; APPLICANT: Cafferkey, Robert
; APPLICANT: Ali, Shujath
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific
; TITLE OF INVENTION: Genes
; FILE REFERENCE: DEX-0239
; CURRENT APPLICATION NUMBER: US/11/230,251
; CURRENT FILING DATE: 2005-09-19
; PRIOR APPLICATION NUMBER: US/09/957,708
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/233,746
; PRIOR FILING DATE: 2000-09-19
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-230-251-26
```

```
Alignment Scores:
Pred. No.: 5,63e-09 Length: 422
Score: 210.50 Matches: 88
Percent Similarity: 39.3% Conservative: 52
Best Local Similarity: 24.7% Mismatches: 109
Query Match: 3.4% Indels: 107
DB: 7 Gaps: 18
```

```
US-10-723-552-3_copy_2307_5741 (1-3435) x US-11-230-251-26 (1-422)
Qy 1982 CAAAAGCTGAGCTCATGGCCCTCACCAAGCTTTGGCGTGGCGAAGGAAATCCATA 1941
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 14 GlnArgAlaGluLeuValAlaValIleThrValLeuGln---AspPheAsnGlnSerIle 32
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 1942 AACATTTATACGGACAGCAGGTATGCTTT---CGGACTGCACACGTACATCGGCGCCATC 1998
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 33 AsnIleValSerAspSerAlaTyrValValGlnAlaThrLysAspIleGluArgAlaLeu 52
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 1999 TATAAACAAAGGGGTTGCTTACCTCAGCAGGGGAGGAATAAAGAACAAAGAG----- 2052
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 53 IleLysTyr-----IleMetAspGlnLeuAsn 62
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2053 GAAATTTCTAAGCCTATTAGAACCGTACATTTACCAAAAGG-----CTAGCTATT 2103
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 63 ProLeuPheAsnLeuLeuGlnGln---AsnValArgLysArgAsnPheProPheTyrIle 81
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2104 ATACACTGCTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAACCCAGATGGCT 2163
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 82 ThrHisIleArgAlaHisThrAsnLeuProGlyProLeuThrArgAlaAsnGluGlnAla 101
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2164 GACCGGTTGCCAAGCAGGAGCCCGAGGTGTTAACTTCTGCTATATAAGAAATGCCC 2223
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 102 AspLeuLeuValSerSerAla----- 108
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2224 AAAGCCCCAGNACCAGACGACAGTACACCTAGAGACTGGCAAGAGATGAAAAAGATA 2283
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 108 ----- 108
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2284 GACCAGTTCTCTGAGACTCCGGAAGGACCTGTATATACCTCAGATGGGAAGAAATCCTG 2343
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 109 -----PheMetGlu----- 111
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2344 CCCACAAAGAGGGTTAGATATGTCCAACAGATACATCGTCTTAACCCAC----- 2394
    | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 112 -----AlaGlnGluLeuHisAlaLeuThrHisValAsnAla 123
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2395 CTAGAACTAAACACACCTGCAGCAGTTGGTCAGAACATCCCTTATCATGTTCTGAGGCTA 2454
    : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 124 IleGlyLeuLysAsnLysPheAspIleThrTrpLysGln----- 136
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2455 CCAGGAGTGGCTGACTCGGTGTCAAACATTGTGTGCCCTGCCAGCTGTTTAAT----- 2508
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 137 -----ThrLysAsnIleValGlnHisCysThrGlnCysGlnIleLeuHisLeuAla 153
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2509 GCTAATCTTCCAGAAATGCTTCCAGGGAAGAGACTAAGGGGAAGCCACCCAGCGCTCAC 2569
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 154 ThrGlnGluAlaArgValAsnPro-----ArgGlyLeuCysProAsnValLeu 169
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2569 TGGGAAGTGGACTTCACTGAGGTAAAGCGGCTAAATACGGGAACAAATACTATTGCTT 2628
    | | | : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 170 TrpGlnMetAspValMetHisVal---ProSerPheGlyLysLeuSerPheValHisVal 188
    | | | : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2629 TTTGTAGACACTTTTTCAGGATGGTAGAGCT---TATCCTACTAAGAAAGAGACTTCA 2685
    | | | : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 189 ThrValAspThrTyrSerHisPheIleTrpAlaThrCysGlnThrGlyGluSerThrSer 208
    | | | : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2686 ACCGTGTGGCTAAAAAATACTGGAAGAAATTTTCCAAGATTTGGAATACCTAAGGTA 2745
    | | | : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 209 HisVal-----LysArgHisLeuLeuSerCysPheProValMetGlyValProGluLys 226
    | | | : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2746 ATAGGTCAGACAATGGTCCAGCTTTTGTCCCGGTAAGTCAGGGACTGGCCCAAGATA 2805
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 227 ValLysThrAspAsnGlyProGlyTyr-----CysSerLysAlaValGlnLysPhe 243
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2806 TTGGGATTGTGAAAA-----CTGATTGTGCATACAGACCCCAAGCTCA 2853
    | | | : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 244 LeuAsn---GlnTrpLysIleThrHisThrIleGlyIleLeuTyrAsnSerGlnGlyGln 262
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Qy 2854 GGACAGGTAGAGAGATGAATAGAACCATTAAGAGACCCCTTACTATAA 2901
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db 263 AlaIleIleGluArgThrAsnArgThrLeuLysAlaGlnLeuValLys 278
    : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
```



Qy	2482	CATTGTGTGCCCTGCCAGCTGGTTAAT-----GCTAATCCTTCCAGAAATGCCTCCAGGG	2533
Db	49	HisCysThrGlnCysGlnIleLeuHisLeuAlaThrGlnGluAlaArgValAsnPro---	67
Qy	2536	AAGAGACTAAGGGGAAGCCACCAGGCGCTCACTGGGAAGTGGACCTTCACTGAGGTAAG	2595
Db	68	-----ArgGlyLeuCysProAsnValLeuTrpGlnMetAspValMetHisVal---	83
Qy	2596	CGGCTAAATACGGAACAATACTATTGCTTTTGTAGACACCTTTTTCAGGATGGGTA	2655
Db	84	ProSerPheGlyLysLeuSerPheValHisValThrValAspThrTyrrSerHisPheIle	103
Qy	2656	GAGGCT---TATCCTACTAAGAAGAGACTTCAACCGTGGCTGAGTAAATAAATACTCGAA	2712
Db	104	TrpAlaThrCysGlnThrGlyGluSerThrSerHisVal-----LysArgHisLeuLeu	121
Qy	2713	GAATTTTTTCCAAGATTTGGAATACCTTAAGGTAAATAGGGTCAGACAAATGGTCCAGCTTTT	2772
Db	122	SerCysPheProValMetGlyValProGluLysValLysThrAspAsnGlyProGlyTyr	141
Qy	2773	GTTGCCAGTAAGTCAGGACTGGCCCAAGATATTTGGGGAATTCATTGGAAA-----	2823
Db	142	CysSerLysAlaPheGln-----LysPheLeuAsn---GlnTrpLysIleThrHis	157
Qy	2824	---CTGCATTGTGCATACAGACACCCCAAGCTCAGGACAGCTAGAGAGGATGAATAGAACC	2880
Db	158	ThrIleGlyIleLeuTyrrAsnSerGlnGlyGlnAlaIleIleGluGlyThrAsnArgThr	177
Qy	2881	ATTAAGAGACCTTACTATAA	2901
Db	178	LeuLysAlaGlnLeuValLys	184

RESULT 11

US-11-175-689-9

Sequence 9, Application US/11175689

Publication No. US20060024806A1

GENERAL INFORMATION:

APPLICANT: STINEAR, TIMOTHY P.

APPLICANT: COLE, STEWART T.

APPLICANT: LEADLAY, PETER F.

APPLICANT: SMALL, PAMELA L.C.

APPLICANT: JOHNSON, PAUL D.R.

APPLICANT: JENKIN, GRANT A.

APPLICANT: DAVIES, JOHN K.

APPLICANT: HAYDOCK, STEPHEN F.

TITLE OF INVENTION: THE MYCOLACTONE LOCUS: AN ASSEMBLY LINE FOR PRODUCING NOVEL POLYKETIDES, THERAPEUTIC AND PROPHYLACTIC USES

FILE REFERENCE: 03495.0329-01

CURRENT APPLICATION NUMBER: US/11/175,689

CURRENT FILING DATE: 2005-07-07

PRIOR APPLICATION NUMBER: 10/987,592

PRIOR FILING DATE: 2004-11-15

PRIOR APPLICATION NUMBER: 60/519,864

PRIOR FILING DATE: 2003-11-14

NUMBER OF SEQ ID NOS: 36

SOFTWARE: PatentIn Ver. 3.3

SEQ ID NO 9

LENGTH: 14130

TYPE: PRT

ORGANISM: Mycobacterium ulcerans

US-11-175-689-9

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

0.00134

146.50

31.0%

19.2%

2.4%

7

14130

243

148

399

473

70

US-10-723-552-3 COPY 2307 5741 (1-3435) x US-11-175-689-9 (1-14130)

QY 40 CGAAGACAGTTGACTTGGAGTGGGACGGGTAAACCCACTCGTTTCTGGTCATACCTGAG 99  
|||||  
Db 7428 ArgArgThrThrArgLeuGlnValSerHisAlaPheHisSer-----ProHis 7443  
QY 100 TGCCGACGACCCCTCTTAGGTAGAGACTTATTGACCAAGATTGGGAGCAGACAAATTTCTTTT 159  
|||||  
Db 7444 MetAsp---ProIleLeu-----GluGlnPheArgGlnIleAlaAlaGlnLeuThrPhe 7460  
QY 160 GAACAGGGAACCCAGAAAGTCTGCGAAATTAACAAACCTATC---ACTGTGTTGACCCCTC 216  
|||||  
Db 7461 -----SerAlaProThrLeuProIleLeuSerAsnLeuThrGly 7473  
QY 217 CAATTAGATGACGAATATCGACTACTCTCCCTAGTAAAGCTGATCAAAATATACAA 276  
|||||  
Db 7474 GlnIleAlaArgHisAspGlnLeuAlaSerPro-----Asp 7485  
QY 277 TTCTGGTTGGAACAG-----TTTCCCAAGCCTGGGCGAGAACCGCA 318  
|||||  
Db 7486 TyrTrpThrGlnGlnLeuArgAsnThrValArgPheHisAspThrValAlaAlaLeuLeu 7505  
QY 319 GGGATGGGTTGGCAAGCAAGTTCCTCCCAAGTATTCACTGAAGCCGACGTGCCACA 378  
|||||  
Db 7506 GlyAlaGly-----GluGlnValPheLeuLeuSerProHis 7518  
QY 379 CCAGTGTCTCAGTCAGACAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTCGCGCGCAT 438  
|||||  
Db 7519 ProVal-----LeuThrGlnAlaIleThrAspThr 7528  
QY 439 GTCCAAAGATTAAATCCACAGGGCATCTCTAGTTCTCT----- 474  
|||||  
Db 7529 ValGluGlnAlaGlyGlyGlyAlaAlaValProAlaLeuArgLysAspArgProAsp 7548  
QY 475 -----GTCCAAATCTCCCTGG 489  
Db 7549 AlaValAlaPheAlaAlaLeuGlyGlnLeuHisCysHisGlyIleSerProSerTrp 7568  
QY 490 AATACT-----CCCTCG---CTACCGGTTAGAAAG----- 516  
|||||  
Db 7569 AsnValLeuTyrCysGlnAlaArgProLeuThrLeuProThrTyrAlaPheGlnHisGln 7588  
QY 517 -----CCTGGGACTAATGACTACTCGACCAGTACAGGACTTGAGAGAGGTC 561  
|||||  
Db 7589 ArgTyrTrpLeuLeuProThrAlaGlyAspPheSer----- 7600  
QY 562 AATAAACGGGTGAGGATATACACCAACAGTCCGAACCCCTTAACCTCTTGTGTGCT 621  
|||||  
Db 7601 GlyAlaAsnThrHisAlaMetHisProLeuLeu-----AspThrAlaThrGlu 7616  
QY 622 CTCCCAACCCCAACGAGCTGGTATACAGTATTGGACTTAAAGGATGCTCTTCTGCTGCTG 681  
|||||  
Db 7617 LeuAlaGluAsnArgGlyTrp-----ValPheThrGly 7627  
QY 682 AGATTACACCCACTAGCCAAACCACTTTTTCCTTCGAATGGAGAGAT----- 729  
|||||  
Db 7628 ArgIleSerProArgThrGlnPro-----TrpLeuAsnGluHisAlaVal 7642  
QY 730 -----CAGGTCAGGGGAAGAACCCGGGAGCTCACCTGGACCCGACTG 771  
|||||  
Db 7643 GluSerAlaValLeuPhePheGlyThrGly----- 7652  
QY 772 CCCCAAGGGTTCAGAACTCCCGGACCATCTTTGACGAAGCCCTACACAGAGACTGGCC 831  
Db 7653 -----PheValGluLeuAlaLeuHis-----ValAla 7661  
QY 832 AACTTCAGGATCCCAACACCCCTCAGGTGACCCCTCCAGTAGTACGTGGATGACCTGCTCTG 891  
|||||  
Db 7662 Asp---ArgAlaGlyTyrSerSerValAsnGluLeuIleValHisThrProLeuLeuLeu 7680  
QY 892 GCGGAGCCCAACACAGGACTGCTTAGAAGCAGCAAGGCACTACTCTCTGGAATGTCT 951  
|||||  
Db 7681 AlaGlyHisAspThrAlaAsp-----LeuGlnIleThrValThrAsp 7696  
QY 952 GACCTAGCTACAGAGCCTCTGCTAAGAAGGCCAGATTTCGAGGAGAGAGGTAAACATAC 1011

Db 7697 AspMetGlyArgGlnSerLeuAsnIleHisSer-----ArgProHisIleGlyHis 7713  
QY 1012 TTGGGGTACAGTTTGGCGGACGGGACAG---CGATGGCTGACGGAGGACGGAAGAAACT 1068  
|||||  
Db 7714 AspAsnThrThrThrGlyAspGluGlnProGluTrpValLeuHisAlaSerAlaValLeu 7733  
QY 1069 GTAGTCCAG-----ATACCGGCCCA----- 1089  
Db 7734 ThrAlaGlnThrThrAspHisAsnHisLeuProLeuThrProValProTrpProPro 7753  
QY 1090 ---ACCAAGCAACAAATGAGAGT---TTGGGGACAGCTGGATTT--- 1134  
Db 7754 GlyThrAlaAlaIleGluValAspPheTyrAspAspLeuAlaAlaGlnGlyTyrAsn 7773  
QY 1135 -----TGCAGACTGTGG-----ATCCCGGG--- 1155  
Db 7774 TyrGlyProThrPheGlnGlyValGlnArgIleTrpArgAspHisAlaThrProAspVal 7793  
QY 1156 ---TTTCGACCTTAGCAGCCCACTCTACCCGCTAACCAAGAAAGGGAATTCCTCC 1212  
Db 7794 IleTyrAlaGluValGluLeuPro-----GluAspThrAspIleAsp 7807  
QY 1213 TGGGCTCTGAGCAGCAGAGGCAATTGATCTATCAAAAGGCGCTCTGAGGCGACCT 1272  
Db 7808 GlyTyrGlyIleHisProAlaLeuPheAspAlaAlaLeuHisProLeuLeuAla----- 7825  
QY 1273 GCTCTGGCCCTCTGAGTAACCTTAAACCTTTTACCCTTTATGTGTGATGAGCGTAAGGA 1332  
Db 7826 -----LeuThrGlnProProThrAsnAspThrAspAspThrAsnThr 7839  
QY 1333 GTAGCCCGGAGTTTAAACCAACCTAGGACCATGGAGAACCTGTGCGCTACCTG 1392  
Db 7840 AlaAspThrGly-----AspGlnValArgLeuProTyrAlaPheThr 7853  
QY 1393 TCAAAAGAAGCTCGATCCTGTAGCCAGTGGTTGGCCCATATCCTTGAAGGCTTATCGCAGCT 1452  
Db 7854 Gly-----IleSerLeuHisAlaThrHisAla 7862  
QY 1453 GTGCCCATATCGTCAAG-----GACCTGACAAATGACTTTTGGGACAGAT 1500  
Db 7863 ThrArgLeuArgValArgLeuThrArgThrGlyAlaAspAlaIleThrValHisThrSer 7882  
QY 1501 ATAACTGTAATAGCCCCCATGCA---TTGGAGAACATCGTTCCGCGACCCCGACCGCA 1557  
Db 7883 AspThrThrGlyAlaProValAlaIleAspSerLeuIleThrArgProLeuThrThr 7902  
QY 1558 TGGATGACCAACGCCGATGACCCACTATCAAAAGCCTTCTCAGAGAGGGTCAAG 1617  
Db 7903 AlaThrGlySerAlaProAlaThrThrAlaAlaGlyLeuLeu-----HisLeuSer 7919  
QY 1618 TTCCTCCACAGCCGCTCTCAACCTGCGCACTCTTCTGCTCGAAGAGACTGTATGAACCA 1677  
Db 7920 TrpProPro-----HisProAspThrThrThrAspThrAspThrAspThrAsp 7935  
QY 1678 GTGACTCATGATGCCATCACTATTGATTGAGGAGACT---GGGGTCCGCAAGGACCTT 1734  
Db 7936 AlaLeuArg-----TyrGlnValIleAlaGluProThrGlnGlnLeuProArgTyrLeu 7953  
QY 1735 ACAGACATACCGCTGAGTGGAGAGTCTAACTTGTCTCACTGACGAGCAAGCAGCTATGTG 1794  
Db 7954 HisAspLeuHisThrSerThrAspLeuHisThrSerThrThrGlu----- 7968  
QY 1795 GTGGAAGGTAGAGGATGGTGGGCGCGGTGTGTGACGGGACCCGACCATCTGGGCC 1854  
Db 7969 -----AlaAspValVal-----ValTrpPro 7975  
QY 1855 AGCAGCTGCGC-----GAAGGAACCTCAGCAAAAGGCTGAGCTCATCGCCTCAGC 1908  
Db 7976 ValProValProSerAsnGluGlnLeuGlnAlaHisGlnAlaSerAspThrAlaValSer 7995  
QY 1909 CAAGCTTGGCGCTGGCCGAAGGGAATCCATAACATTTATACGGAC----- 1956  
|||||

Db 7996 SerArgIleHisThrLeuThrArgGlnThrLeuThrValValGlnAspTrpLeuThrHis 8015  
Qy 1957 -----AGCAGGTATGCTTTCGCACTGCACAC----- 1983  
Db 8016 ProAspThrThrGlyThrArgLeuValIleValThrArgHisGlyValSerThrSerAla 8035  
Qy 1984 -----GTACATGGGCGCATCTATAAACAAGGGGTTGCTTACC 2022  
Db 8036 HisAspProValProAspLeuAlaHisAlaValTrp-----GlyLeuIleArg 8052  
Qy 2023 TCACAGGAGGGAATAAAGACAAAGAGGAATTTCTA----- 2061  
Db 8053 SerAlaGlnAsnGluHisProGlyArgPheThrLeuLeuAspThrAspAsnThrAsn 8072  
Qy 2062 -----AGCCTATTAGAACCGTACATTTACCAAAAGG-----CTAGCTATT--- 2103  
Db 8073 SerAspThrLeuThrAlaLeuThrLeuProThrArgGluAsnGlnLeuAlaIleArg 8092  
Qy 2104 -----ATACACTGTCTGGACATCAGAAAGCTAAAGATCTCATATCCAGAGGAAC 2154  
Db 8093 ArgAspThrIleHisIlePro----- 8099  
Qy 2155 CAGATGGCTGACCGGTTGCCAAGCAGCAGCCAGGGTGTAACTTCTGCTATAATA 2214  
Db 8100 -----ArgLeuThrArgHisSerSerAspGlyAlaLeuThrAlaProValVal 8115  
Qy 2215 GAAATGCCAAAGCCCAAGACCCAGACAGACAGTACACCCCTAGAAGACTGGCAAGAGATA 2274  
Db 8115 ----- 8115  
Qy 2275 AAAAGATAGACCACTTCTTGAGACTCCGGAAGGAGCTGCTATACCTCAGATGGGAAG 2334  
Db 8116 -----ValAsp-----ProGluGlyThrValLeuIleThrGlyGlyThr 8128  
Qy 2335 GAAATCTGCCCCACAAAGAGGGTTAGATATATGCCAAGACATACATCGTCTAACCCAC 2394  
Db 8129 GlyThrLeu-----GlyAlaLeuPheAlaGlu-----HisLeuValSerAla 8142  
Qy 2395 CTAGGAATAAACACCTGCAGCAGTGTGTCAGAACATCCCTTATCATGTTCTCAGAGCTA 2454  
Db 8143 HisGlyValArgHisLeuLeuLeuThrSerArgGlyPro-----GlnAla 8158  
Qy 2455 CCAGAGTGGTGTGACTCGGTGGTCAAAACATTTGTGTGCCCTGCCAGCTGGTTAATGCTAAT 2514  
Db 8159 HisGlyAlaThrAsp----- 8163  
Qy 2515 CCTTCCAGATGCTCCAGGGAAGACACTAAGGGGAAGCCACCAGCGCTCACTGGGAA 2574  
Db 8164 -----LeuGlnGlnArgLeuThrAspLeuGlyAlaHisValThr 8176  
Qy 2575 GTGACTTCTACTGAGTAAAGCCCGCTAAATACGGAACAAATACCTATTGTTGTTTGTGA 2634  
Db 8177 IleThrAlaCysAspIleSerAspPro----- 8185  
Qy 2635 GACACCTTTTCAGATGGTAGAGGCTTATCTTACTAAGAAAGAGACTTCAACCGTGGTG 2694  
Db 8186 GluAlaLeuAlaLeuValAsnSerValProThrGlnHisArgLeuThrAlaValVal 8205  
Qy 2695 GCTAAAAAATACCTGGAAGAAATTTTCCAGATTTTGGATACCTAAGGTAAATAGGTCA 2754  
Db 8206 HisThrAlaAlaValLeuAlaAspThrPro-----ValThrGluLeuThrGlyAsp 8222  
Qy 2755 GACAATGGTCCAGCTTTTGTGTCAGGTAAGTCAAGGACTGGCCAGAGATATTGGGGATT 2814  
Db 8223 -----GlnLeuAspGlnValLeuAlaProLysIleAspAla 8234  
Qy 2815 GATTGGAACTGCATTGTGCATACAGACCCCAAGAGCTCAGACAGCTAGAGAGATGAAT 2874  
Db 8235 AlaTrpGlnLeuHis----- 8239  
Qy 2875 AGAACCAATTAAAGAGACCCCTTACTAATTGACCGCGAGAGACTGGCGTTAATGATTGGATA 2934  
Db 8240 -----GlnLeuThrTyrGluHisAsn----- 8246

Qy 2935 GCTCTCTGCGCTTTTGTGCTTTTATAG-----GTTAGGAACACCCCTGGACAG 2982  
Db 8247 ---LeuSerAlaPheIleMetPheSerSerMetAlaGlyMetIleGlySerProGlyGln 8265  
Qy 2983 -----TTT 2985  
Db 8266 GlyAsnTyrAlaAlaAlaAsnThrAlaLeuAspAlaLeuAlaAspTyrArgHisArgLeu 8285  
Qy 2986 GGGCTGACCCCTATGAATTACTCTACGG-----GGACCCCCCCCA 3027  
Db 8286 GlyLeuProAlaThrSerLeuAlaTrpGlyTyrTrpGlnThrHisThrGly----- 8302  
Qy 3028 TTGCTAGAAATGCTCTGTACATAGTGTGACGTGCTTCCAGCCTTTGCCAGCCTTTGTTCTCT 3087  
Db 8303 -----LeuThrAlaHisLeuThrAspValAspLeuAlaArg-----MetThr 8316  
Qy 3088 AGGCTCAAGCAGCTTTCAGATCCCATCGTTTCCAGTGGGAGGCACTCCCGGAGGCTTAC 3147  
Db 8317 ArgLeuGlyLeuMetProIleAlaThrSerHisGlyLeuAlaLeuPheAspAlaAlaLeu 8336  
Qy 3148 TCAGGAGGAGGAGACTTTCAGATCCCATCGTTTCCAGTGGGAGGATTCAGTCTACGTT 3207  
Db 8337 AlaThrGlyGlnProValSerIlePro-----AlaProIleAsnThrHisThrLeuAla 8354  
Qy 3208 AGACCCACCGCTGCAGGAAACCTCGAGACTCGGTGGAAGGCGCTTATCTC---GTACTT 3264  
Db 8355 ArgHisAlaArgAspAsnThrLeu-----AlaProIleLeuSerAlaLeu 8369  
Qy 3265 TTGACCAACACA-----ACGGCTGTGAAAGTCGAA 3294  
Db 8370 IleThrThrProArgArgAlaAlaSerAlaAlaThrAspLeuAlaAlaArgLeuAsn 8389  
Qy 3295 GGAATCTCC 3303  
Db 8390 GlyLeuSer 8392

## RESULT 12

US-11-175-689-7  
; Sequence 7, Application US/11175689  
; Publication No. US20080024806A1  
; GENERAL INFORMATION:  
; APPLICANT: STINEAR, TIMOTHY P.  
; APPLICANT: COLE, STEWART T.  
; APPLICANT: LEADLAY, PETER F.  
; APPLICANT: SMALL, PAMELA L.C.  
; APPLICANT: JOHNSON, PAUL D.R.  
; APPLICANT: JENKIN, GRANT A.  
; APPLICANT: DAVIES, JOHN K.  
; APPLICANT: HAYDOCK, STEPHEN F.  
; TITLE OF INVENTION: THE MYCOLACTONE LOCUS: AN ASSEMBLY LINE FOR PRODUCING  
; TITLE OF INVENTION: NOVEL POLYKETIDES, THERAPEUTIC AND PROPHYLACTIC USES  
; FILE REFERENCE: 03495.0329-01  
; CURRENT APPLICATION NUMBER: US/11/175,689  
; PRIOR FILING DATE: 2005-07-07  
; PRIOR APPLICATION NUMBER: 10/987,592  
; PRIOR FILING DATE: 2004-11-15  
; PRIOR APPLICATION NUMBER: 60/519,864  
; PRIOR FILING DATE: 2003-11-14  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 3.3  
; SEQ ID NO 7  
; LENGTH: 16990  
; TYPE: PRT  
; ORGANISM: Mycobacterium ulcerans  
US-11-175-689-7

Alignment Scores:  
Pred. No.: 0.00241 Length: 16990  
Score: 143.50 Matches: 242  
Percent Similarity: 31.4% Conservative: 151  
Best Local Similarity: 19.3% Mismatches: 409  
Query Match: 2.3% Indels: 451

DB:	7	Gaps:	69
US-10-723-552-3_COPY_2307_5741 (1-3435) x US-11-175-689-7 (1-16990)			
Qy	40	CGAAGAACAGTTGACTTTGGAGTGGGACGGGTAAACCCACTCGTCTTCTGGTCATCACTCGAG	99
Db	15848	ArgArgThrThrArgLeuGlnValSerHisAlaPheHisSer-----ProHis	15863
Qy	100	TGCCAGACACCCCTCTTAGGTAGAGACTATTATGACCAAGATGGGAGCACAATTTCTTTT	159
Db	15864	MetAsp-----ProIleLeu-----GluGlnPheArgGlnIleAlaAlaGlnLeuThrPhe	15880
Qy	160	GAACAAGGGAACACAGAGTCTGCAAAATACAAACCTATC---ACTGTCTGACCCCTC	216
Db	15881	-----SerAlaProThrLeuProIleLeuSerAsnLeuThrGly	15893
Qy	217	CAATTAGATGCAATATCGACTACTCTCCCTCTAGTAAAGCCTGATCAAAATATACAA	276
Db	15894	GlnIleAlaArgHisAspGlnLeuAlaSerPro-----Asp	15905
Qy	277	TTCTGGTTGGACAG-----TTTCCCAAGCCTGGGCGAGAACCGCA	318
Db	15906	TyrTrpThrGlnGlnLeuArgAsnThrValArgPheHisAspThrValAlaAlaLeuLeu	15925
Qy	319	GGGATGGCTTTGGCAAGCAAGTTCCTCCCAAGTATTCAACTGAAGGCCAGTGCCACA	378
Db	15926	GlyAlaGly-----GluGlnValPheLeuGluLeuSerProHis	15938
Qy	379	CCAGTGTCACTCAGACAGTACCCCTTGAGTAAAGAGCTCAAGAGGAATTCGGCCGCAT	438
Db	15939	ProVal-----LeuThrGlnAlaIleThrAspThr	15948
Qy	439	GTCCAAGATTAATCCACAGGGCATCTAGTTCTCT	474
Db	15949	ValGluGlnAlaGlyGlyGlyAlaAlaValProAlaLeuArgLysArgProAsp	15968
Qy	475	-----GTCCAATCTCCCTGG	489
Db	15969	AlaValAlaPheAlaAlaLeuGlyGlnLeuHisCysHisGlyIleSerProSerTrp	15988
Qy	490	AATACT-----CCCCTG---CTACCGGTTAGAAAG-	516
Db	15989	AsnValLeuTyrCysGlnAlaArgProLeuThrLeuProThrThrAlaPheGlnHisGln	16008
Qy	517	-----CCTGGGACTAATGACTATCGACCATCGACAGTACAGGACTTGAGAGAGTGC	561
Db	16009	ArgTyrTrpLeuLeuProThrAlaGlyAspPheSer-----	16020
Qy	562	AATAAAGGGTGCAGGATATACACCAACAGTCCCGAACCCCTTATAACCTCTTGTGCT	621
Db	16021	GlyAlaAsnThrHisAlaWechHisProLeuLeu-----AspThrAlaThrGlu	16036
Qy	622	CTCCACACCCCAACGGAGCTGGTATACAGTATTGGACTTAAAGGATGCGCTCTTCTGCGCTG	681
Db	16037	LeuAlaGluAsnArgGlyTrp-----ValPheThrGly	16047
Qy	682	AGATTACACCCCACTAGCCACCACTTTTGGCTTCGAATGGAGAGATCCAGGTACGGGA	741
Db	16048	ArgIleSerProArgThrGlnPro-----TrpLeuAsnGluHisAlaVal	16062
Qy	742	AGACCGGGCAGCTCACCTGGACCGGACTGCCCAAGGTTCAAGAACTCCCGACCATC	801
Db	16063	GlusErAlaValLeu-----PheProAsnThrGlyPheVal	16074
Qy	802	TTTGACGAAGCTTACACAGAGACCTGGCCAACTTCAGGATCCACACCCCTCAGGTGACC	861
Db	16075	---GluLeuAlaLeuHis-----ValAlaAsp---ArgAlaGlyTyrSerSerValAsn	16090
Qy	862	CTCTCCAGTAGTGGATGACTGCTTCTGGGGAGGCCCAACACAGGACTGCTTAGAA	921
Db	16091	GluLeuIleValHisThrProLeuLeuLeuAlaGlyHisAspThrAlaAsp	16107
Qy	922	GGCAGGAAGGCACTACTGCTGGAATTGCTGACCTAGGCTACAGAGCCTCTGCTAAGAAG	981
Db	16108	---LeuGlnIleThrValThrAspThrAspMetGlyArgGlnSerLeuAsnIleHis	16126
Qy	982	GCCAGATTTTCAGGAGAGGTAACATACTTGGGGTACAGTTTGGCGGACGGCAG---	1038
Db	16127	Ser-----ArgProHisIleGlyHisAspAsnThrThrThrGlyAspGluGlnPro	16143
Qy	1039	CGATGGCTGACGGAGCAGCGAAGAAAACGTAGTCCAG-----	1077
Db	16144	GluTrpValLeuHisAlaSerAlaValLeuThrAlaGlnThrThrAspHisAsnHisLeu	16163
Qy	1078	-----ATACCGGCCCA-----ACCACAGCCAAACAAATAGAGAGTTT	1116
Db	16164	ProLeuThrProValProTrpProProGlyThrAlaAlaIleGluValAspAspPhe	16183
Qy	1117	-----TTGGGGACAGCTGGATT-----TGCGAGA	1140
Db	16184	TyrAspAspLeuAlaAlaGlnGlyTyrAsnTyrGlyProThrPheGlnGlyValGlnArg	16203
Qy	1141	CTGTGG-----ATCCCGGG-----TTTGGACCTTAGCAGCCCCACTCTAC	1182
Db	16204	IleTrpArgAspHisAlaThrProAspValIleTyrAlaGluValGluLeuPro-----	16221
Qy	1183	CCGTAAACCAAGAAAGGGAATTCTCTGGGCTCTCTGAGCACACAGAGCGCATTTGAT	1242
Db	16222	-----GluAspThrAspIleAspGlyTyrGlyIleHisProAlaLeuPheAsp	16237
Qy	1243	GCTATCAAAAGCCCTGCTGAGCGACCTGCTCTGGCCCTCCCTGAGTAACATAACCC	1302
Db	16238	AlaAlaLeuHisProLeuLeuAla-----LeuThrGlnPro	16249
Qy	1303	TTTACCCCTTATGTGGATGAGCGTAAGGAGTAGCCCGGGAGTTTAAACCAAAACCTA	1362
Db	16250	ProThrAsnAspThrAspAspThrAsnThrAlaAspThrGly-----	16263
Qy	1363	GGACCATGGAGAACACCTGTGCTCTACCTGTCAAAAGAGCTCGATCTGTAGCCAGTGGT	1422
Db	16264	AspGlnValArgLeuProTyrAlaPheThrGly-----	16274
Qy	1423	TGGCCCATATGCTGAGAGGCTATCGAGCTGTGCCATCTACTGTCACAG-----	1470
Db	16275	-----IleSerLeuHisAlaThrHisAlaThrArgLeuArgValArgLeuThrArgThr	16292
Qy	1471	GACCTCACAAATGACTTTGGGACAGAATACTGTAAATAGCCCCCATGCA---TTG	1527
Db	16293	GlyAlaAspAlaIleThrValHisThrSerAspThrThrGlyAlaProValAlaIleIle	16312
Qy	1528	GAGAACATCGTTCCGCGACGCCACCGATGGATGACCAACGCCCGCATGCCACTAT	1587
Db	16313	AspSerLeuIleThrArgProLeuThrThrAlaThrGlySerAlaProAlaThrThrAla	16332
Qy	1588	CAAGCCTGCTTCTCACAGAGAGGTACGTTGCTCCACAGCGGCTCTCAACCTGCC	1647
Db	16333	AlaGlyLeuLeu-----HisLeuSerTrpPro-----HisProAsp	16345
Qy	1648	ACTCTTCTGCTGAAGAGACTGATGAACACGACTCATGATGTCATCACTATTGATT	1707
Db	16346	ThrThrThrAspThrAspThrAspThrAspThrAspAlaLeuArgTyrGlnValIleAla	16365
Qy	1708	GAGGAGACT---GGGGTCCGCAAGACCTTACAGACATACCGCTGCTGGAGAAGTGCTA	1764
Db	16366	GluProThrGlnGlnLeuProArgTyrLeuHisAspLeuHisThrSerThrAspLeuHis	16385
Qy	1765	ACCTGTTTCACTGACGGAGCAGCTATGTGTGGAAGTAAGAGTAGGCTGGGGCGCG	1824
Db	16386	ThrSerThrThrGlu-----AlaAsp	16392
Qy	1825	GTGTGTGACGGGACCGCACGATCTGGCCACGAGCCTGCGC-----GAAGAACTTCA	1878
Db	16393	ValVal-----ValTrpProValProValProSerAsnGlnGluLeuGln	16407
Qy	1879	GCACAAAGGCTGAGCTCATGGCCCTCACGCAAGCTTTGCGGCTGGCCGAAGGAAATCC	1938

Db 16408 AlaHisGlnAlaSerAspThrAlaValSerSerArgIleHisThrLeuThrArgGlnThr 16427  
Qy 1939 ATAAACATTATACCGAC-----AGCAGGTATGCTTT 1971  
Db 16428 LeuThrValGlnAspThrLeuThrHisProAspThrThrGlyThrArgLeuValIle 16447  
Qy 1972 GCGACTGCACAC-----GTACATGGG 1992  
Db 16448 ValThrArgHisGlyValSerThrSerAlaHisAspProValProAspLeuAlaHisAla 16467  
Qy 1993 GCCATCTATAAACAAAGGGTGTCTTACCTCAGCAGGAGGGAATAAAGCAACAGAG 2052  
Db 16468 AlaValTrp-----GlyLeuIleArgSerAlaGlnAsnGluHisProGlyArgPhe 16484  
Qy 2053 GAAATTCTTA-----AGCTATTAGAACCGCTACATTTA 2085  
Db 16485 ThrLeuLeuAspThrAspAsnThrAsnSerAspThrLeuThrAlaLeuThrLeu 16504  
Qy 2086 CCAAAAGG-----CTAGCTATT-----ATACACTGTCCTGGACATCAG 2124  
Db 16505 ProThrArgGluAsnGlnLeuAlaIleArgArgAspThrIleHisIlePro----- 16521  
Qy 2125 AAGCTAAAGATCTCATATCCAGAGAAACAGATGGCTGACCGGTGGCCAGCAGCA 2184  
Db 16522 -----ArgLeuThrArgHisSer 16527  
Qy 2185 GCCCAGGTGTAACTTCTGCTATATAGAAATGCCAAAGCCAGAACCCAGACGA 2244  
Db 16528 SerAspGlyAlaLeuThrAlaProVal----- 16537  
Qy 2245 CAGTACACCTAGAAAGCTGCAAGAGATAAAAGATAGACAGTCTCTGAGACTCCG 2304  
Db 16538 -----ValAsp-----Pro 16540  
Qy 2305 GAAGGACCTGCTATACCTCAGATGGGAGGAATCTGCGCCCAAGAAAGGGTTAGAA 2364  
Db 16541 GluGlyThrValLeuIleThrGlyGlyThrGlyThrLeu-----GlyAlaLeu 16556  
Qy 2365 TATGTCACAGATACATCGTCTAACCCACCTAGAACATAACACCTGAGCAGCTGGTC 2424  
Db 16557 PheAlaGlu-----HisLeuValSerAlaHisGlyValArgHisLeuLeuThrSer 16574  
Qy 2425 AGAACATCCCTTATCATGTTCTCAGGCTACAGAGTGGCTGACTCGGTGGTCAACAT 2484  
Db 16575 ArgArgGlyPro-----GlnAlaHisGlyAlaThrAsp----- 16585  
Qy 2485 TGTGTCCTGCCAGCTGTTAATGCTAATCTTCCAGATGCTCCAGGAGACGACTA 2544  
Db 16586 -----LeuGlnGln 16588  
Qy 2545 AGGGGAAGCCACCCAGCGCTCACTGGGAAGTGACTTCACTAGGTAAAGCCGCTAAA 2604  
Db 16589 ArgLeuThrAspLeuGlyAlaHisValThrIleThrAlaCysAspIleSerAspPro--- 16607  
Qy 2605 TACGGAAACAAATACCTATTGGTTTTTGTATAGACCTTTTCAGGATGGGTAGAGCTTAT 2664  
Db 16608 -----GluAlaLeuAlaAlaLeuValAsnSerVal 16617  
Qy 2665 CCTACTAAGAAGAGACTTCAACCGTGGTGGCTAAAAAATACTGGGAAGAAATTTTCCA 2724  
Db 16618 ProThrGlnHisArgLeuThrAlaValValHisThrAlaAlaValLeuAlaAspThrPro 16637  
Qy 2725 AGATTGGATACCTAAGGTAAATAGGTGCAGACAATGGTCCAGCTTTTGTGGCCAGGTA 2784  
Db 16638 -----ValThrGluLeuThrGlyAsp-----GlnLeu 16646  
Qy 2785 AGTCAGGAGCTGCCAAGATATTGGGATGATTGGAACACTGCATTTGTGCATACAGACCC 2844  
Db 16647 AspGlnValLeuAlaProLysIleAlaAspAlaAlaIleArgGlnLeuHis----- 16661  
Qy 2845 CAAAGCTCAGGACAGGTAGAGAGGATGAATAGAACCACTTAAGAGACCCCTTACTAAATTG 2904  
Db 16662 -----GlnLeu 16663

Qy 2905 ACCCGGAGACTGGCTTAATGATTGGATAGCTCTCTGCTCCCTTTGTCTTTTAGG--- 2961  
Db 16664 ThrTyrGluHisAsn-----LeuSerAlaPheIleMetPheSerSer 16677  
Qy 2962 -----GTTAGAACACCCCTGGACAG----- 2982  
Db 16678 MetAlaGlyMetIleGlySerProGlyGlnGlyAsnTyrAlaAlaAlaAsnThrAlaLeu 16697  
Qy 2983 -----TTTGGGTGAGCCCTATGAATTACTCTACCGG 3015  
Db 16698 AspAlaLeuAlaAspTyrArgHisArgLeuGlyLeuProAlaThrSerLeuAlaIleArg 16717  
Qy 3016 -----GGACCCCTCCCATTTGTAGAAATTTGCTTGTACATAGTCTCT 3057  
Db 16718 TyrTrpGlnThrHisThrGly-----LeuThrAlaHisLeuThr 16730  
Qy 3058 GACGTGCTGCTTTCCAGCCTTTCTCTAGGCTCAAGGACCTTGAGTGGGTGAGACAA 3117  
Db 16731 AspValAspLeuAlaArg-----MetThrArgLeuGlyLeuMetProIleAlaThrSer 16748  
Qy 3118 CGACGTGGAGGCAACTCCGGGAGGCTACTCAGGAGGAGGAGACTTGCAGATCCCAT 3177  
Db 16749 HisGlyLeuAlaLeuPheAspAlaAlaLeuAlaThrGlyGlnProValSerIlePro--- 16767  
Qy 3178 CGTTTCAAGTGGGAGATTCTAGTCTAGTACGCCACCGTCAGGAAACCTCGAGACT 3237  
Db 16768 ---AlaProIleAsnThrHisThrLeuAlaArgHisAlaArgAspAsnThrLeu----- 16784  
Qy 3238 CGGTGGAGGCGCTTATCTC---GTACTTTTGACACACCA----- 3276  
Db 16785 -----AlaProIleLeuSerAlaLeuIleThrProArgArgAlaAlaSer 16801  
Qy 3277 -----ACGGCTGTGAAAGTCGAAGGATCTCC 3303  
Db 16802 AlaAlaThrAspLeuAlaAlaArgLeuAsnGlyLeuSer 16814

RESULT 13  
US-11-143-980-47  
; Sequence 47, Application US/11143980  
; Publication No. US20050272133A1  
; GENERAL INFORMATION:  
; APPLICANT: Hucul, John  
; APPLICANT: Helti, Bradley A.  
; APPLICANT: Wagonaar, Melissa M.  
; APPLICANT: Graziani, Edmund  
; APPLICANT: Summers, Mia  
; APPLICANT: Kulowski, Kerry  
; APPLICANT: Pong, Kevin  
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Production of a Complex  
; TITLE OF INVENTION: Polyketide  
; FILE REFERENCE: AM-101426US  
; CURRENT APPLICATION NUMBER: US/11/143,980  
; CURRENT FILING DATE: 2005-06-03  
; PRIOR APPLICATION NUMBER: US 60/664,483  
; PRIOR FILING DATE: 2005-03-23  
; PRIOR APPLICATION NUMBER: US 60/576,895  
; PRIOR FILING DATE: 2004-06-03  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 47  
; LENGTH: 5712  
; TYPE: PRF  
; ORGANISM: Streptomyces sp.  
US-11-143-980-47

Alignment Scores:  
Pred. No.: 0.0394 Length: 5712  
Score: 126.50 Matches: 259  
Percent Similarity: 29.8% Conservative: 150  
Best Local Similarity: 18.9% Mismatches: 455  
Query Match: 2.0% Indels: 511



Db 1440 -----ValAspTrpSerGlyGlyArgValGluLeu-----LeuThr 1451  
Qy 1753 GGAGAGTGCTAACCTGTTCACTACGGAAGCAGCTATGTGTGGAAGGTAAAGAGGATG 1812  
Db 1452 Gly---ThrThrProTrpProThrGly-----GlyLeuArgArg 1464  
Qy 1813 GCTGGGGGGGGTG-----GTGACCGGACCCGACG-----ATCTGGGCCAGCAGGCTG 1863  
Db 1465 AlaGlyValSerSerPheGlyValSerGlyThrAsnAlaHisValIleLeuGluGlnVal 1484  
Qy 1864 CCGGAA-----GGAACTTCACGACCAAAAG 1887  
Db 1485 ProGluThrAlaArgProThrGlyProIleGlyGluAspAspGlyGluAlaAlaProVal 1504  
Qy 1888 GCTGAGCTCATGGCC-----CTCAGCCAGCTTTCGGCTGGCC 1926  
Db 1505 AlaTrpValLeuSerGlyGlnGlyGluThrGlyLeuArgAlaGlnAlaGluArgLeuCys 1524  
Qy 1927 GAAGGGAATCCATAAATTTAT---ACGACAGCAGGTATGCTTTGCGACTGACAC 1983  
Db 1525 AlaPheMetAlaAlaAspThrArgProThrProAlaGluValGlyTrpSerLeuAlaSer 1544  
Qy 1984 GTACATGGGGCCATCTATAACAAAGGGGTGCTTACCTCAGCAGGAGGGAATAAAG 2043  
Db 1545 ThrArg---AlaThrLeuSerHisArgAlaValValValGlyAlaGlyArg----- 1560  
Qy 2044 AACAAAGAGAAATTTCTAAGCCTATTAGAAGCCGTACATTTACCAAAAAGGCTAGCTATT 2103  
Db 1561 -----AspGluLeuLeuArgGlyValAsnAlaVal----- 1570  
Qy 2104 ATACACTGCTCGACATCAGAAAGCTAAAGATCTCATATCCAGAGAAACAGATGGCT 2163  
Db 1570 ----- 1570  
Qy 2164 GACCGGTTGCCAAGCAGCCAGCGGTGTTAACTTCTGCTCTATAATAGAAATGCC 2223  
Db 1571 -----AlaAsnGlyThrProValProGlyValVal----- 1580  
Qy 2224 AAAGCCCCAGAACCCAGCAGCAGTACACCCCTAGAAGACTGGCAAGAGATAAAAAAGATA 2283  
Db 1580 ----- 1580  
Qy 2284 GACCAGTTCTGTAGACTCCGGAAGGACCTGCTATACCTCAGATGGAAGGAATCCTG 2343  
Db 1581 -----ArgGlyThrGlyAlaSerGlyAspValValPheValPhe 1593  
Qy 2344 CCCACAAAGAGGTTAGATATGTTCCACAGATACATCGTCTAACCCACTAGGAAT 2403  
Db 1594 Pro---GlyGlnGlySerGlnTrpValGlyMet----- 1603  
Qy 2404 AACACCTGCAGCAGTGGTGAGAACATCCCTTATCATGTTCTGAGGCTACACGAGTG 2463  
Db 1604 -----AlaLeuGluLeuValGluSerSerProValPheAlaArgArgLeuGlyAspCys 1621  
Qy 2464 GCTGACTCGGTG----- 2475  
Db 1622 AlaAspAlaLeuAlaProPheValGluTrpSerLeuPheAspValLeuGlyAspGluVal 1641  
Qy 2476 -----GTCAACATTGTGCTCCCTGCCAGCTGTTAATGCTA----- 2512  
Db 1642 AlaIleGlyArgVal-AspValValGlnProValLeuTrpAlaValMetValSerLeuAl 1661  
Qy 2513 -----ATCCTTCCAGATGCTCCAGGGAAGAGACTAAAGGGA 2550  
Db 1661 aGluLeuTrpArgSerPheGlyValValProSerAlaValValGlyHisSerGlnGlyG 1681  
Qy 2551 A---GCCACCCAGCGCTCACTGGGAAGTGGACTTCACTGAGTAAAGCCGCTAAATAC 2607  
Db 1681 uileAlaAlaCysValAlaGlyAlaLeuThr----- 1692  
Qy 2608 GGAACAAATACCTATTGGTTTTGTAGACACCTTTTCAGATGGGTAGAGCTTATCT 2667  
Db 1693 -----LeuGluAspGlyAlaArgValValAl 1701

Qy 2668 ACTAAGAAAGAGACT-----TCAACCGTGGTGGCTAAAAA 2702  
Db 1701 aLeuArgSerArgAlaLeuLeuAlaLeuSerGlyArgGlyGlyMetValSerValProVa 1721  
Qy 2703 AATACTGGAAGAAATTTTCCAGATTTTGAATACCTTAAGTAATAGGTCAGACAATGG 2762  
Db 1721 lSerAlaAspArgLeuArgAspArgValGlyLeuSer-----ValAlaAlaValAsnGl 1739  
Qy 2763 TCCAGCTTTT-----GTTGCCAGGTAAGTCAGGACTG----- 2796  
Db 1739 yProAlaSerThrValValSerGlyAlaValGluValLeuGluAlaValLeuAlaGluPh 1759  
Qy 2797 -----GCCAAGATATTGGGATTGTTGAAACTGCATTGTGCATACAGACCCCAAG 2849  
Db 1759 eProGluAlaAlaArgIleProValAspTyrAlaSerHis----- 1772  
Qy 2850 CTCAGCAGCTAGCAGAGGATGAATAGAACCATTAAGAGACCTTACTTAATTTGACCGC 2909  
Db 1773 -SerValGlnValGluGlyIleArgGluGlyLeuAlaGluAlaLeuAlaProValArgPr 1792  
Qy 2910 GGAGACTGGC-----GTTAATGATTGGATAGC 2936  
Db 1792 oArgThrGlyGlnValProPheTyrSerThrValThrGlyArgLeuMetAspThrIleGl 1812  
Qy 2937 TCTCTGCTCCCTTGTGCTTTTGGGTTAGGAACACCCCTGGACAGCTTTGGCTGACCCC 2996  
Db 1812 uLeuAspAlaGluTyrTrpTyrArgAsnLeuArgGluThrValGluPheGlnSerThrVa 1832  
Qy 2997 CTATGAATTTACTCTACGGGGA-----CCCCCCCCATTGGT 3032  
Db 1832 lGluHisLeuMetArgGlnGlyHisThrValPheValGluAlaSerProHisProValLe 1852  
Qy 3033 AGAAAT-----GCTTCTGTACATAGTGTGCTGCTGCTTTTCCCAGCC 3077  
Db 1852 uThrIleGlyValGlnAspThrAlaAspThrThrAspThrAspIleValValThr----- 1870  
Qy 3078 TTTGTTCTCTAGCTCAAGGCACCTTGTAGTGGTGAGACAACGAGCGTGGAGGCAACTCCG 3137  
Db 1871 -----GlySerLeuArgArgAspAspGlyThrValGlnArgPheLeuThrSerLeuAl 1888  
Qy 3138 GGAG-----GCCTACTCAGAGAGGAGAGA 3161  
Db 1888 aGluLeuHisValArgGlyValArgIleAspTrpGlyProLeuPheAlaGlyValSerPr 1908  
Qy 3162 CTTGCAGATCCCA---CATGTTTCCAGTGGGAGATTCACTCTACGTTAGACCCACCG 3218  
Db 1908 oValGluLeuProThrTyrAlaPheGln----- 1917  
Qy 3219 TGCAGGAAACCTCGAGACTCGGTGGAAGGCGCTTATCTGCTACTTTTGACCACCAAC 3278  
Db 1918 -----ArgGluArgPheTrpLeuGly-----Al 1925  
Qy 3279 GGTGTGAAAGTCAAGGAATCTCCACTCGATCCATGATCCACCTTAACCGCGGCC 3338  
Db 1925 aAspIleAlaGluSerAlaValAspThrTrpArgTyrGlnIleSerTrpLysProLeuPr 1945  
Qy 3339 ACCTCCCGAT-----TCGGGG---TGGAAAGCC 3363  
Db 1945 oAspMetAspProProAlaLeuSerGlyThrTrpLeuAla 1958

## RESULT 14

US-10-995-561-659  
; Sequence 659, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24





## RESULT 15

```
US-11-109-156-14
; Sequence 14, Application US/11109156
; Publication No. US20050250144A1
; GENERAL INFORMATION:
; APPLICANT: Toshio Ota
; APPLICANT: Takao Isogai
; APPLICANT: Tetsuo Nishikawa
; APPLICANT: Koji Hayashi
; APPLICANT: Kaoru Otsuka
; APPLICANT: Jun-ichi Yamamoto
; APPLICANT: Shizuko Ishii
; APPLICANT: Tomoyasu Sugiyama
; APPLICANT: Ai Wakamatsu
; APPLICANT: Keiichi Nagai
; APPLICANT: Tetsuji Otsuki
; APPLICANT: Shin-ichi Funahashi
; APPLICANT: Chiaki Senoo
; APPLICANT: Jun-ichi Nezu
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEIN KINASE/PROTEIN
; TITLE OF INVENTION: PHOSPHATASE
; FILE REFERENCE: 06501-099002
; CURRENT APPLICATION NUMBER: US/11/109,156
; CURRENT FILING DATE: 2005-04-19
; PRIOR APPLICATION NUMBER: US/10/060,065
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: PCT/JP00/05061
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: US 60/183,322
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: JP 11-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-183767
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: JP 2000-241899
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-109-156-14

Alignment Scores:
Pred. No.: 0.0262 Length: 745
Score: 126.00 Matches: 134
Percent Similarity: 33.5% Conservative: 77
Best Local Similarity: 21.3% Mismatches: 219
Query Match: 2.0% Indels: 200
DB: 7 Gaps: 32

US-10-723-552-3_COPY_2307_5741 (1-3435) x US-11-109-156-14 (1-745)
QY 805 GACGAACCCCTACACAGACCTGGCCAACTTCAGGATCCACACCTCAGGTGACCCCTC 864
DB 138 AsnLysIleIleHisArgAspLeu-----LysProGluAsnIleVal 151
QY 865 CTCAGTACGTGGATGACCTGCTTCTGGCGGAGCCACCAACAGGACTGTTAGAGGC 924
DB 152 LeuGlnAspVal-----Gly 156
QY 925 ACGAAGCACTACTGCTGGAATTGCTGACTAGGCTACAGACCTCTGTAAGAAGGCC 984
DB 157 GlyLys---IleIleHisIleIleAspLeuGlyTyrAlaLysAspValAspGlnGly 175
QY 985 CAGATTGACGAGGAGGTAAACATCTTGGGGTACAGTTTGGGGACGGGACGCGATGG 1044
DB 176 SerLeuCysThrSerPheValGlyThrLeuGlnTyrLeuAlaProGlu----- 191
```

```
QY 1045 CTGACGGAGGACGACGAAGAAAACCTGTAGTCCAGATACCGGCCCAACACACGCAACAA 1104
DB 192 LeuPheGluAsnLys-----ProTyrThrAlaThrVal 202
QY 1105 ATGAGAGAGTTTTTGGGGACAGCTGGATT---TGCAGACTGTGGATCCCGGGTTTCGG 1161
DB 203 AspTyrTrpSerPheGlyThrMetValPheGluCys-----IleAlaGlyTyrArg 219
QY 1162 ACCTTAGCAGCCCACTCTACCCGCTAACCAAGAAAAGGGGAATCTCTCGGCTCCT 1221
DB 220 ProPheLeuHisHisLeuGlnPro-----PheThrTrp----- 230
QY 1222 GAGCACCAGAAGGCATTGTGCTATCAAAAAGCCCTGTGTGAGCGCACCTGCTCTGGCC 1281
DB 231 ---HisGluLys-----IleLysLys----- 236
QY 1282 CTCCTGACGTAACAAACCTTTTACCCCTTTATGTG---GATGAGCGTAGGGAGTAGGCC 1338
DB 237 -----LysAspProLysCysIlePheAlaCysGluGluMetSerGlyGluVal 252
QY 1339 CGG-----GGAGTTTAAACCAACCCCTA 1362
DB 253 ArgPheSerSerHisLeuProGlnProAsnSerLeuCysSerLeuIleValGluProMet 272
QY 1363 GGCACCATGGAGAAGACCTGTCGCCTACCTGTCAAAGAAGCTCGCATCTGTAGCCAGTGGT 1422
DB 273 GluAsnTrpLeuGln-----LeuMetLeuAsnTrpAspProGlnGlnArgGly 288
QY 1423 TGGCCCATATGCTGGAAGGCTATCGCAGCTGTGGCCATATCGTCAAGGACGCTGACAAA 1482
DB 289 GlyProVal-----Asp 292
QY 1483 TTGACTTTTGGACAG---AATATAACTGTAATAGCCCCCATCATTCATGGAGAACATCGTT 1539
DB 293 LeuThrLeuLysGlnProArgCysPheValLeuMetAspHisIleLeu----- 308
QY 1540 CGGCAGCCCCCAGACCGATGGATGACCAACGCCCGCATGACCCCACTATCAAAAGCTGTT 1599
DB 309 -----AsnLeuLysIleValHisIleLeuAsnMetThr 319
QY 1600 CTCACAGAGAGGTGCTGCTTCACACAGCGCTCTCAACCTCGCCACTCTTCTGCT 1659
DB 320 SerAlaLysIleIleSerPheLeuLeuProAspGluSerLeuHisSerLeuGlnSer 339
QY 1660 GAAGAGACTGATGAACCCAGTCACTCATGATTGCCATCAACTATTATTGATTGAGGAGCTGG 1719
DB 340 ArgIleGluArgGluThrGlyIleAsnThrGlySerGlnGluLeuLeuSerGluThrGly 359
QY 1720 GTC-----CGCAAGGACCTTACAGACATACCGCTGACTGGA----- 1755
DB 360 IleSerLeuAspProArgLysProAlaSerGlnCysValLeuAspGlyValArgGlyCys 379
QY 1756 ---GAAGTGTAACCTGGTTCACTGACGGAAGCAGCTATGTGTGGAAAGGTAAAGAGATG 1812
DB 380 AspSerTyrMetValTyrLeuPheAspLysSerLysThrValTyrGluGlyPro---Phe 398
QY 1813 GCTGGGGCGCGGTGGTGAC----- 1833
DB 399 AlaSerArgSerLeuSerAspCysValAsnTyrIleValGlnAspSerLysIleGlnLeu 418
QY 1834 -----GGGACCCGACGATCTGGCC-----AGCAGCTCCCG 1866
DB 419 ProIleIleGlnLeuArgLysValTyrAlaGluAlaValHisTyrValSerGlyLeuLys 438
QY 1867 GAAGGAACCTTCACACAAAGAGGTGAGCTCATGGCCCTCACGCAAGCTTTGGCGCTGCC 1926
DB 439 GluAspTyrSer-----ArgLeuPhe 445
QY 1927 GAAGGGAATCCATAAACATTTATACGACAGCAGGTATGCCTTTGGCAGCTGCACACGTA 1986
DB 446 GlnGlyGlnArgAlaAlaMetLeuSerLeuLeuArgTyr----- 458
```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 15:55:36 ; Search time 9.89233 Seconds  
(without alignments)  
3199.272 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_5620\_7533  
Perfect score: 3468  
Sequence: 1 ATGCATCCACGTTAAACCG.....CTAGCAGGAGAGCTGGCCGC 1914

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0  
Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSSWEB.spool/US10723552/runat 14022006 125146 13079/app query.fasta\_1  
-DB=Issued Patents AA -QPMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abss02p -USER=US10723552 @CGN 1.1 161 @runat 14022006 125146 13079  
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5 COMB.pcp:\*  
2: /cgn2\_6/ptodata/1/iaa/6 COMB.pcp:\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pcp:\*  
4: /cgn2\_6/ptodata/1/iaa/PCRTUS\_COMB.pcp:\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pcp:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3420	98.6	638	2	US-09-376-781-6
2	2866.5	82.7	660	2	US-09-111-085-2
3	2866.5	82.7	660	2	US-09-376-781-5
4	2422	69.8	657	2	US-09-111-085-4
5	2422	69.8	657	2	US-09-376-781-4
6	2395	69.1	657	2	US-09-171-553B-10
7	2353	67.8	656	2	US-09-376-781-3
8	2353	67.8	656	2	US-09-171-553B-6
9	1489.5	42.9	667	2	US-09-315-127-5
10	1489.5	42.9	667	2	US-09-315-127-6
11	1489.5	42.9	667	2	US-09-070-630-13
12	1477.5	42.6	673	2	US-09-075-272-5

13	1319	38.0	665	2	US-09-309-572-14	Sequence 14, Appl
14	1319	38.0	665	2	US-09-718-096-14	Sequence 14, Appl
15	1319	38.0	1312	2	US-09-554-572-26	Sequence 26, Appl
16	1318.5	38.0	632	2	US-09-315-127-2	Sequence 2, Appl
17	1318.5	38.0	632	2	US-09-315-127-3	Sequence 3, Appl
18	1277	36.8	645	2	US-09-315-127-8	Sequence 8, Appl
19	1277	36.8	645	2	US-09-315-127-9	Sequence 9, Appl
20	1262.5	36.4	654	2	US-09-315-127-11	Sequence 11, Appl
21	1262.5	36.4	654	2	US-09-315-127-12	Sequence 12, Appl
22	692.5	20.0	469	1	US-08-484-126-1	Sequence 1, Appl
23	692.5	20.0	469	1	US-09-374-909-1	Sequence 1, Appl
24	659	19.0	445	1	US-08-447-925-6	Sequence 3, Appl
25	639.5	18.4	453	1	US-08-484-126-3	Sequence 3, Appl
26	639.5	18.4	453	2	US-09-374-909-3	Sequence 7, Appl
27	626.5	18.1	196	1	US-08-484-126-7	Sequence 7, Appl
28	626.5	18.1	196	2	US-09-374-909-7	Sequence 7, Appl
29	448	12.9	567	1	US-08-007-282B-2	Sequence 2, Appl
30	414.5	12.0	538	2	US-09-175-328-4	Sequence 4, Appl
31	413.5	11.9	540	2	US-09-949-016-11511	Sequence 11511, A
32	413.5	11.9	540	2	US-09-719-554-26	Sequence 26, Appl
33	413.5	11.9	685	2	US-09-719-554-35	Sequence 35, Appl
34	369	10.6	493	2	US-08-979-847B-106	Sequence 106, App
35	356.5	10.3	192	2	US-08-486-099-107	Sequence 107, App
36	356.5	10.3	192	2	US-08-360-107A-117	Sequence 117, App
37	356.5	10.3	192	2	US-08-484-223B-107	Sequence 107, App
38	356.5	10.3	192	2	US-08-919-597-107	Sequence 107, App
39	356.5	10.3	192	2	US-08-475-668A-107	Sequence 107, App
40	356.5	10.3	192	2	US-08-485-551A-107	Sequence 107, App
41	356.5	10.3	192	2	US-08-471-913A-107	Sequence 107, App
42	356.5	10.3	192	2	US-08-485-264A-107	Sequence 107, App
43	356.5	10.3	192	2	US-08-474-349A-107	Sequence 107, App
44	356.5	10.3	192	2	US-08-470-896-107	Sequence 107, App
45	356.5	10.3	192	2	US-08-485-546A-107	Sequence 107, App

ALIGNMENTS

RESULT  
US-09-376-781-6  
; Sequence 6, Application US/09376781  
; Patent No. 6261806  
; GENERAL INFORMATION:  
; APPLICANT: Banerjee, Papi T.  
; APPLICANT: Patience, Clive  
; APPLICANT: Andersson, Goran K.  
; TITLE OF INVENTION: Molecular Sequences of Swine Retrovirus and Methods of  
; Patent No. 6261806  
; TITLE OF INVENTION: Use  
; FILE REFERENCE: 61750-267  
; CURRENT APPLICATION NUMBER: US/09/376,781  
; CURRENT FILING DATE: 1999-08-18  
; EARLIER APPLICATION NUMBER: 60/097,015  
; EARLIER FILING DATE: 1998-08-18  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PERV-C  
; OTHER INFORMATION: polypeptide sequence taken from GenBank Accession  
; OTHER INFORMATION: No. 6261806 AF038600 for comparison.  
US-09-376-781-6

Alignment Scores:  
Pred. No.: 0 Length: 638  
Score: 3420.00 Matches: 637  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.8% Mismatches: 0  
Query Match: 98.6% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-09-376-781-6 (1-638)

```
QY 1 ATGCATCCACGTTAAACCGGGCCACCTCCGATTCGGGTGGAAACCCGAAAGACTG 60
Db 1 MethHisProThrLeuAsnArgHisLeuProIleArgGlyGlyProLysArgLeu 20
QY 61 AAAATCCCTTAAGCTTCGCTCCATCGCTGGTTCCTTACTCTGCTCAATAACCTCTCAG 120
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40
QY 121 ACTAATGGTATCGCATAGGACAGCCTGAACTCCCATAAACCTTATCTCTCACCTGG 180
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTACTGACTCGGCACAGGTATTAATATCAACACACTCAAGGGAGGCTCCTTTA 240
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTCAGATCAGTTATTCCTAGTCTGACCTCA 300
Db 81 GlyThrTrpTrpProAspLeuTrpValCysLeuArgSerValIleProSerLeuThrSer 100
QY 301 CCCCCAGATATCCTCCATCGCTCACGGATTTTATGTTTCCCGCAGGACCAACCAATAATGGA 360
Db 101 ProProAspIleLeuHisAlaHisGlyPheTrpValCysProGlyProProAsnAsnGly 120
QY 361 AAACATTCGGGAATCCGAGAGATTTCTTTGTGTAACCAATGGAATGTTAACTCTTAAT 420
Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140
QY 421 GATGATATTCGAATGCGCAACCTCTCAGCAGGATAGGTAAGTTTCTTATGTCTCAAC 480
Db 141 AspGlyTrpTrpLysTrpProThrSerGlnAspArgValSerPheSerTrpValAsn 160
QY 481 ACCTATACCGAGCTCTGGCAATTAATTAACCTGACCTCGATTAAGTTCCTGATCTCAAC 540
Db 161 ThrTrpThrSerSerGlyGlnPheAsnTrpLeuThrTrpIleArgThrGlySerProLys 180
QY 541 TGCTCTCTTCAGACTAGATTACCTAAATAATAGTTTCACTGGAAGAAGGAAACAAGAA 600
Db 181 CysSerProSerAspLeuAspTrpLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200
QY 601 AATATCTTAAATGGTAAGTATGTCCTGGGAATGGTATATTATGGAGGCTCGGGT 660
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTrpTrpGlySerGly 220
QY 661 AAACAACGAGCTCCATTCCTAATCTCGCTCAAAATAAACAGCTGGAGCTCCAATG 720
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240
QY 721 GCTATAGACCAATACGGTCTTGACGGGTCAAGACCCCAACCCCAAGGACCGAGCCA 780
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260
QY 781 TCCTCTACATAACTTCCTGGATCAGACCCCACTAGTCTTAACAGCACCACTAAATGGG 840
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerSerSerThrThrLysMetGly 280
QY 841 GCAAAACCTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAG 900
Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300
QY 901 GCTACCTCTTCTGTGTAGCTATGCTTAGCTCGGGCCCACTTACTATGAAGGAATGGCT 960
Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTrpTrpGlyMetAla 320
QY 961 AGAAGGGAATTCATGTGAAGAAGACATAGACCAATGCATGCATGGGATGCCAA 1020
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340
QY 1021 AATAAGCTTACCTTACTGAGTTTCTGGAAGAGGCACCTGCATAGGAAGGTCCCCCA 1080
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360
```

```
QY 1081 TCCCAACCAACCTTTGTAAACCACTGAAGCCCTTTAATCAAACTCTGAGAGTCAATAT 1140
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTrp 380
QY 1141 CTGTACTCTGTTATGACAGAGTGGGCATGTAACTACTGTGATTAAACCCCTTGTGTTC 1200
Db 381 LeuValProGlyTrpAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400
QY 1201 ACCTTGGTTTTTAACCAAACTAAAGATTTTTCGATTATGTCCTCAAAATGTTCCCGAGTG 1260
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420
QY 1261 TATTACTATCCGAAAAGCAATCCTTGATGAAATATGACTACAGAAATCATCGCAAAAG 1320
Db 421 TyrTrpTrpProGluLysAlaIleLeuAspGluTrpAspTrpArgAsnHisArgGlnLys 440
QY 1321 AGAGAACCCATATCTCTGACACTTGTGATGTCGAGCTTGGAGTGGCAGCAGGTGTA 1380
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460
QY 1381 GGAACAGAAACAGCTGCCCTGGTCACGGACCAACAGAGCTAGAGAACAGGACTTAGTAAC 1440
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
QY 1441 CTACATCGAATTGTAAACAGAGATCTCCAGCCCTAGAAAATCTGTCAAGTAACCTGGAG 1500
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuLysSerValSerAsnLeuGlu 500
QY 1501 GAATCCCTAACTCTTATCTGAAGTAGTCTCAGAAATAGAGAGGCTTAGATTTATTA 1560
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520
QY 1561 TTTCTAAAGAGAGGATTATGTAGCTTGAAGGAGGAATGCTGTTTATGTGGAT 1620
Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTrpValAsp 540
QY 1621 CATTCAGGGCCATCAGAGACTCCATGACNAAGCTTAGAGAAAGGTTGGAGAGGCTCGA 1680
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560
QY 1681 AGGAAAAGGAAACTACTCAAGGGTGGTTTCAGGGATGGTTCAACAGGCTCTCTTGGTTG 1740
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeu 580
QY 1741 GCTACCTTACTTTCTGCTTAAACAGGACCTTAATAGTCTCTCTCTCTTACTACAGTT 1800
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600
QY 1801 GGGCCATCTATTATTAAACAGTTAATTCCTTCATTAGAGAACGAATAAGTGCAGTCCAG 1860
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620
QY 1861 ATCATGTACTTAGACAACAGTACCAAGCCCGCTCTAGCAGGAGAGCTGGCCGC 1914
Db 621 IleMetValLeuArgGlnGlnTrpGlnSerProSerArgGluAlaGlyArg 638
```

## RESULT 2

```
US-09-111-085-2
; Sequence 2, Application US/09111085
; Patent No. 610034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
```



```
Db      641 GlnIleMetValLeuArgGlnGlnTyrGln 650
RESULT 3
US-09-376-781-5
; Sequence 5, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papiya T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PERV-A
; OTHER INFORMATION: polypeptide sequence taken from GenBank Accession
; OTHER INFORMATION: No. 6261806 Y12238 for comparison.
US-09-376-781-5

Alignment Scores:
Pred. No.:      7, 728-281      Length:      660
Score:          2866.50        Matches:    537
Percent Similarity: 88.6%      Conservative: 39
Best Local Similarity: 82.6%   Mismatches:  53
Query Match:     82.7%        Indels:      21
DB:              2            Gaps:         6

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-09-376-781-5 (1-660)
QY      1  ATGCATCCACGTTAAACGGCGCCACCTCCCGATTTCGGGGTGGAAAGCCGAAAGACTG 60
Db      1  MethHisProThrLeuSerArgArgHisLeuProIleArgGlyGlyProLysArgLeu 20
QY      61  AAAATCCCTTAAAGCTTCGCTCCATCGCGGTTCCTTACTCTCTGTCATTAACCTCTCAG 120
Db      21  LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
QY      121 ACTAATGGTATGCGCATAGGAGACAGCCTGAACTCCCATAAACCTTATCTCTCACCTGG 180
Db      41  ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
QY      181 TTAATTACTGACTCCGGCACAGTATTAAATATCAACAACACTCAAGGGAGGCTCTTTTA 240
Db      61  LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
QY      241 GGAACCTGGTGGCTGATCTATACGTTGCTCCTCAGATCAGTATTCTCTAGTCTG----- 294
Db      81  GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY      295 ---ACCTCACCCCCAGATATCTCCATGCTCAGCGATTATTGTTTGCCAGGACCACCA 351
Db      101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
QY      352 AATAATGAAACAATTGGGGAATCCCGAGATTTCTTTTGTGTAACAATGGAACTGTGTA 411
Db      121 AsnAsnGluGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysIle 140
QY      412 ACCTCTAATGATGATATATGGAAATGGCCAACTCTCAGCAGGATAGGGTAAGTTTCT 471
Db      141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTrpSer 160
QY      472 TATGTCAACACCTATACCAGCTCTCGACAAATTTAATTAC-----CTGACC 516
Db      161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
QY      517 TGG-----ATTAGAACTCGAAGCCCCAAGTGTCTCTCTTCAGAC 555
Db      181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
QY      556 CTAGATTACTTAAATAAGTTTCACCTCAGAAAGGAAACAAAGAAATATCTTAAATGG 615
Db      201 LeuAspTyrLeuLysIleSerPheThrGlnLysGlyGlnGlnAsnIleGlnLysTrp 220
QY      616 GTAATGTATGTCTTGGGGAATGATATATATGAGGCTCGGTAAACAACACCGGCTCC 675
Db      221 ValAsnGlyIleSerTrpGlyIleValTyrTyrGlySerGlyArgLysLysGlySer 240
QY      676 ATTCTAATCTTCGCTCAAAATA---AACAGCTGGAGCTTCCAATGGCTATAGGACCA 732
Db      241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProProValAlaIleGlyPro 260
QY      733 AATACGGTCTTGACGGGTCAAAGACCCCAACCCAA-----GGACCAAGGACCA 780
Db      261 AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGlnArgProSerProAsnPro 280
QY      781 TCCTCT---AACATAAATCTTCTGGATCAGACCCCACTGAGTCTAAACAGCACACTAAATG 837
Db      281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300
QY      838 GGGGCAAACTTTTTCAGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCACCAGCTCCA 897
Db      301 GlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320
QY      898 GAGGCTACCTCTTCTTGTGGCTATGCTTACGCTTCGGGCCCACTTACTATGAGGAATG 957
Db      321 GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGlnGlyMet 340
QY      958 GCTAGAGAGGAGAAATTCATGTGACAAAGAAACATAGAGACCAATGACACATGGGATCC 1017
Db      341 AlaArgGlyGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySer 360
QY      1018 CAAATAAGCTTACCTTACTGAGGTTCCTGGAAAGGCACCTGTCATAGGAAGGTTCC 1077
Db      361 GlnAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyMetValPro 380
QY      1078 CCATCCCAACCAACCTTTGTAACCACTCAAGCTTTTAATCAACCTCTGAGAGTCAA 1137
Db      381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400
QY      1138 TATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTATATCTGGATTAAACCCCTTGT 1197
Db      401 TyrLeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVal 420
QY      1198 TCCACTTGGTGGTTTAAACCAACTAAGATTTTTCATTTGCTGATTTGTTCCCGCA 1257
Db      421 SerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArg 440
QY      1258 GTGTATTACTATCCGAAAGCAATCCTTCATGAATATGACTACAGAAATCATCGACAA 1317
Db      441 ValTyrTyrProGluLysAlaValLeuAspGluTyrAspTyrArgTyrAsnArgPro 460
QY      1318 AAGAGAACCCATATCTCTGACACTTCTGTCGATGCTCGGACTTGGAGTGGCAGCAGGT 1377
Db      461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGly 480
QY      1378 GTAGAAACAGAACAGCTGCCCTGTCACGGGACCAAGCAGCAGCTAGAACAGGACTTAGT 1437
Db      481 ValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSer 500
QY      1438 AACCTACATCCAAATGTAAACAGAAAGACTCTCAAGCCCTAGAAAATCTGTAGTAACCTG 1497
Db      501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520
QY      1498 GAGGAATCCCTTAAACCTCTTATCTGAAGTAGTCTCTACAGAAATAGAAAGGGGTAGATT 1557
```



Db 521 GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 540  
 Qy 1558 TTATTTCTAAAGAGGAGGATTATGTAGCTTTGAAAGGAGGAATGCTGTTTTATGTG 1617  
 Db 541 LeuPheLeuLysGluGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrVal 560  
 Qy 1618 GATCATTCAGGGCCATCAGAGACTCCATGAAACAGCTTAGAGAAAGTTGAGAGCGT 1677  
 Db 561 AspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArg 580  
 Qy 1678 CGAAGGGAAAGAAACTACTCAAGGGTGGTTTCAGGGATGGTTCAACAGGCTCTTTGG 1737  
 Db 581 ArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrPheAsnArgSerProTyr 600  
 Qy 1738 TTGGCTACCTTCTTCTGCTTTTAAACAGGACCTTAAATAGTCTCTCTCTGTTACTACA 1797  
 Db 601 MetThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThr 620  
 Qy 1798 GTTGGGCCATGTATTTAAACAAGTTAATTCCTTCATTAGAACAGATAAGTGCAGTC 1857  
 Db 621 ValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaVal 640  
 Qy 1858 CAGATCATGTGTTAGACACAGTACCAC 1887  
 Db 641 GlnIleMetValLeuArgGlnGlnTyrGln 650  
 RESULT 4  
 US-09-111-085-4  
 ; Sequence 4, Application US/09111085  
 ; Patent No. 6100034  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stoye, Jonathan P  
 ; APPLICANT: Weiss, Robin A  
 ; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope  
 ; TITLE OF INVENTION: specific sequences  
 ; FILE REFERENCE: 4238/75168  
 ; CURRENT APPLICATION NUMBER: US/09/111.085  
 ; CURRENT FILING DATE: 1998-07-07  
 ; EARLIER FILING DATE: 1997-05-16  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 657  
 ; TYPE: PRT  
 ; ORGANISM: Porcine retrovirus  
 US-09-111-085-4  
 Alignment Scores:  
 Pred. No.: 7,14e-236 Length: 657  
 Score: 2422.00 Matches: 457  
 Percent Similarity: 80.2% Conservative: 68  
 Best Local Similarity: 69.8% Mismatches: 96  
 Query Match: 69.8% Indels: 34  
 DB: 2 Gaps: 9  
 US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-09-111-085-4 (1-657)  
 Qy 1 ATGCATCCACAGTTAAACCGCGCCACTCCCGATTCCGGGTGAAAGCCGAAAGACTG 60  
 Db 1 MethisProthrLeuSerTyrArgHisLeuProthrArgGlyGlyGluProLysArgLeu 20  
 Qy 61 AAAATCCCTTAAGCTTCGCTCCATCGCTGGTGTCTTACTCTGTCAATAACCTCTCAG 120  
 Db 21 ArgIleProLeuSerPheAlaSerIleAlaTyrPheLeuThrLeuThrIleThrProGln 40  
 Qy 121 ACTAATGGTATGGCATAGGAGACAGCTCGAATCCCAATAAACCTTATCTCACTCGG 180  
 Db 41 AlaSerLysArgLeuLeuAspSerSerAsnProHisArgProLeuSerLeuThrTyr 60  
 Qy 181 TTAATTAAGTACCTCCGCGACAGGTATTAATTAATCAACAACTCAAGGGAGGCTCCTTTA 240  
 Db 61 LeuIleIleAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProArg 80

Qy 241 GGAACTGGTGGCTGATCTATACGCTTTCCTCAGATCAGTTATTTCTAGTCTGACCTCA 300  
 Db 81 GlyThrTyrTrpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100  
 Qy 301 ---CCCCAGATATCTCCATGCTCAGGATTTTATGTTTCCAGGAGCACCAATAAT 357  
 Db 101 ThrProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGlu 120  
 Qy 358 GGAAAAATTCGCGAAATCCAGAGATTTCTTTGTAACAATGGAACCTGTGTAACCTCT 417  
 Db 121 ---LysTyrCysGlyGlySerGlyGluSerPheCysArgArgTyrPheCysValThrSer 139  
 Qy 418 AATGATGGATATTTGAAATGGCAACCTCTCAGCAGGATAGGTAAGTTTTTCTTATGTC 477  
 Db 140 AsnAspGlyAspTyrLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159  
 Qy 478 AACACCTATACAGCTCTCGGACAAATTAATTAACCTGACCTGGATTAGAAGTGGAGCCCC 537  
 Db 160 Asn-----SerGlyProGlyLysTyrLysValMetLysLeuTyrLysAspLys 175  
 Qy 538 AAGTGCTCTCTTCAGACCTAGATTACCTAAATAAGTTTCTACCTGAGAAAGGAAACAA 597  
 Db 176 SerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGln 195  
 Qy 598 GAAAAATATCTAAATGGTAAATGGTATGTCTTGGGGAATGGTATAT-----TATGGA 651  
 Db 196 GluAsnIleGlnLysTyrIleAsnGlyMetSerTrpGlyIleValPheTyrLysTyrGly 215  
 Qy 652 GGCTCGGGTAAACAAACAGGCTCCATTCTAACTATTCGCCTCAAAATAAACACAG---CTG 708  
 Db 216 GlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 232  
 Qy 709 GAGCTCCATGCTATAGGACCAATACCGTCTTGACGGGTCAAGACCCCAACCCAA 768  
 Db 233 GluProProValAlaValGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 252  
 Qy 769 GGACCA-----GGA 777  
 Db 253 GluProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrGln 272  
 Qy 778 CCATCTCTTAACATAACTTCTGGATCAGACCCCACTGAGTCT-----AACAGC----- 825  
 Db 273 ProProSerAsnGlyThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 292  
 Qy 826 ---ACGACTAAATGGGGCAAACTTTTATGCTCATCCAGGAGCTTTTCAAGCTCTT 882  
 Db 293 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 312  
 Qy 883 AACTCCACGACTCCAGAGGCTACTCTTCTTGTGGTATGCTTAGCTTCGGGCGCCACCT 942  
 Db 313 AsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProPro 332  
 Qy 943 TACTATGAAGAAATGGCTAGAAGAGGAAATCAATGTGACAAAGAACATAGAGACAA 1002  
 Db 333 TyrTyrGluGlyMetAlaLysGluGlyLysPheAsnValThrLysGluHisArgAsnGln 352  
 Qy 1003 TGCACATGGGGATCCCAAAATAAGTCTTACTGAGGTTTCTGAAAGGACCTGTC 1062  
 Db 353 CysThrTrpGlySerArgAsnLysLeuThrGluValSerGlyLysGlyThrCys 372  
 Qy 1063 ATAGGAAGGTTCCCCCATCCCAACACCTTTGTAACCACTGAAGCTTAAATCAA 1122  
 Db 373 IleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 392  
 Qy 1123 ACCTCTGAGAGTCAATATCTGGTACCTGGTTATGACAGGTGGTGGGCATGTAATACCTGGA 1182  
 Db 393 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGly 412  
 Qy 1183 TTAACCCCTTGTTTCCACCTTGGTTTTTAAACCAACTAAAGATTTTTCATTATGCTGTC 1242  
 Db 413 LeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheCysValMetVal 432

```
QY 1243 CAAATTGTTCCCGAGTGATTACTATATCCCGAAAGCAATCCTTGATGAATATGACTAC 1302
Db 433 GlnlleValProArgValTyrHisProGluGluValValLeuAspGluTyrAspTyr 452
QY 1303 AGAAATCATGACAAAAGAGAACCCCATATCTCTGACACTTGTGTGATGCTCGGACTT 1362
Db 453 ArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeu 472
QY 1363 GGAGTGGCAGAGGTGTAGGAACAGGAACAGCTGCCCTGGTCACGGGACCAACAGAGTAA 1422
Db 473 GlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 492
QY 1423 GAAACAGACTTAGTAACCTCATCGAATTGTAAACAGAAAGTCTCCAAAGCCCTAGAAAAA 1482
Db 493 GlnLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuGluGlu 512
QY 1483 TCTGTGAGTAACCTGGAGGAATCCCTAACCTCCTTATCTGAAGTAGTCTCTACAGATAGA 1542
Db 513 SerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArg 532
QY 1543 AGAGGTTAGATTTATTATTTCTAAAGAGAGGATTATGTGTAGCCCTGAAGAGGAA 1602
Db 533 ArgGlyLeuAspLeuLeuPheLeuArgGluGlyLeuCysAlaAlaLeuLysGluGlu 552
QY 1603 TGCTGTTTATGTGATCATTACGGGGCCATCAGAGACTCCATCAACAAGCTTAGAGAA 1662
Db 553 CysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGlu 572
QY 1663 AGTTTGGAGAGCGCTCGAAGGGAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGGTTC 1722
Db 573 ArgLeuGluArgArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrPhe 592
QY 1723 AACAGGCTCTTGTGTGGTACCTACTTCTGCTTTAAACAGAGCCCTTAATAGTCCTC 1782
Db 593 AsnArgSerProTyrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu 612
QY 1783 CTCCTGTTTACTCACAGTTGGGCCATGTATTATTAAACAAGTTAATTGCTTCATTAGAGAA 1842
Db 613 LeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGlu 632
QY 1843 CGAATAAGTCAGTCACGATCATGTCTTACTTAGACAAACAGTAGTACCAA 1887
Db 633 ArgValSerAlaValGlnIleMetValLeuArgGlnTyrGln 647
```

## RESULT 5

```
US-09-376-781-4
; Sequence 4, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papija T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; EARLIER FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:PERV-B
; OTHER INFORMATION: polypeptide sequence taken for comparison from
; OTHER INFORMATION: GenBank Accession No. 6261806 Y12239.
US-09-376-781-4
```

## Alignment Scores:

```
Pred. No.: 7.14e-236 Length: 657
Score: 2422.00 Matches: 457
Percent Similarity: 80.2% Conservative: 68
Best Local Similarity: 69.8% Mismatches: 96
Query Match: 69.8% Indels: 34
DB: 2 Gaps: 9

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-09-376-781-4 (1-657)

QY 1 ATGCATCCCGAGTTAAACCGCGCCACCTCCCGATTGGGGTGAAGCCGAAAGACTG 60
Db 1 MethisProThrLeuSerTrpArgHisLeuProThrArgGlyGlyGluProLysArgLeu 20
QY 61 AAAATCCCTTAAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTGTAATACCTCTCAG 120
Db 21 ArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrIleThrProGln 40
QY 121 ACTAATGTATGTCGCATAGAGACAGCTCGAATCCCATAAACCTTATCTCTCACCTGG 180
Db 41 AlaSerSerLysArgLeuIleAspSerSerAsnProHisArgProLeuSerLeuThrTrp 60
QY 181 TTAATTAATGACTCCGCGACAGGTATTAAATATCAACAACTCAAGGGAGGCTCCTTTA 240
Db 61 LeuIleIleAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProArg 80
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTCAGATCAGTTATTCTTAGTCTGACCTCA 300
Db 81 GlyThrTrpTrpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
QY 301 ---CCCCAGATATCTCCATCTCAGCGATTTTATGTTGCCCGAGGACCAACAAATAAT 357
Db 101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGlu 120
QY 358 GGAATAACATTCGGAATAATCCAGAGATTTCTTTGTAAACAATGGAACCTGTGTACCTCT 417
Db 121 ---LysTyrCysGlyGlySerGlyGluSerPheCysArgArgTyrPserCysValThrSer 139
QY 418 AATCATGATATTCGGAATGCGCAACCTCTCAGCAGGATAGGTAAGTTTCTTATATGTC 477
Db 140 AsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
QY 478 AACACCTATACAGCTCTGGCAATTTAATATACCTGACCTGGATTAGAACCTGGAGCCCC 537
Db 160 Asn-----SerGlyProGlyLysTyrLysValMetLysLeuTyrLysAspLys 175
QY 538 AAGTGTCTCTTCAGACCTAGATTACCTAAAAATAAGTTTCACCTGAGAAAGGAAAAACAA 597
Db 176 SerCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGln 195
QY 598 GAAATATCTTAAATGGTAAATGGTATGCTTTGGGAATGGTATAT-----TATCGA 651
Db 196 GluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTyrLysTyrGly 215
QY 652 GCCTCGGTAACCAACACAGGCTCCATCTACTATTCGCTCAAAATAAACACAG---CTG 708
Db 216 GlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 232
QY 709 GAGCTCTCAATGGTATAGGACCAAAATACGGCTTTGACGGGTCAAAGACCCCAACCCAA 768
Db 233 GluProProValAlaValGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 252
QY 769 GGACCA-----GGA 777
Db 253 GluProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrGln 272
QY 778 CCATCCTCTTAACATAACTTCTGGATCAGACCCCACTGAGTCT-----AACAGC----- 825
Db 273 ProProSerAsnGlyThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 292
QY 826 ---ACGACTAAATGGGGGCAAAACTTTTAGCCTCATCCAGGAGCTTTTCAAGCTCTT 882
Db 293 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 312
```

```
QY 883 AACTCCAGGCTCCAGAGGCTACCTCTCTTTGGTGTATGCTTAGCTTCGGGCCACCT 942
Db 313 AmsSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProPro 332
QY 943 TACTATGAAGAAAGGCTAGAGAGGAAATTCATGTGACAAAGACACAGACCA 1002
Db 333 TyrTrpGluGlyMetAlaLysGluGlyPheAsnValThrLysGluHisArgAsnGln 352
QY 1003 TGCACATGGGGATCCCAAAATAAGCTTACCCCTTACTGAGGTTCCTGGAAGGACCTGC 1062
Db 353 CysThrTrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCys 372
QY 1063 ATAGAAAGGTTCCCCCATCCACACACCTTTGTAAACACACTGAAGCCCTTAAATCAA 1122
Db 373 IleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValThrGluGln 392
QY 1123 ACCTCTGAGAGTCAATCTCGTACCTGGTTATCACAGGTGGTGGCATGTAATCTGGA 1182
Db 393 AlaSerGluAsnGlnTrpLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGly 412
QY 1183 TTAACCCCTTGTTTCCACCTTGGTFTTTTAAACCAAACTAAAGATTTTTCATTATGTC 1242
Db 413 LeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheCysValMetVal 432
QY 1243 CAATTTGTTCCCGAGTGATTACTATCCGAAAAAGCAATCTTGTATGATATGACTAC 1302
Db 433 GlnIleValProArgValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyr 452
QY 1303 AGAAATCATGACAAAAGAGAGACCCATATCTCTGACACTGCTGTGATGCTCGGACTT 1362
Db 453 ArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeu 472
QY 1363 GGAGTGGCAGCAGGTGTAGGAACAGGAACAGCTCCCTGGTCCAGGACACACAGCTA 1422
Db 473 GlyThrAlaValGlyValGlyThrGlyThrAlaLeuLeuThrGlyProGlnGlnLeu 492
QY 1423 GAAACAGGACTTAGTAACCTACATCGAATGTGTAAACAGAGATCTCCAGGCCCTAGAAAA 1482
Db 493 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuGluGlu 512
QY 1483 TCTGTAGTAACCTGGAGGATCCCTAACCTCTTATCTGAGTGTAGTCTACAGAAATAGA 1542
Db 513 SerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArg 532
QY 1543 AGAGGTTAGATTTATTTATTTCTAAAGAGGAGGATTTATGTGTAGCTTTGAAGGAGAA 1602
Db 533 ArgGlyLeuAspLeuLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGlu 552
QY 1603 TGCTGTTTTATGTGATCATTTCAAGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAA 1662
Db 553 CysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGlu 572
QY 1663 AGTTGGAGAAGCGTCCAGGGAAGAAAGAACTACTCAAGGTGGTTTGAAGGATGCTTC 1722
Db 573 ArgLeuGluArgArgArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPhe 592
QY 1723 AACAGGTCTCTTGTGGTGTACCTACTTCTTCTTAAACAGGACCTTAAATAGTCTTC 1782
Db 593 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu 612
QY 1783 CTCTGTGTACTCAGTGTGGCCATGATTTATTAAACAAGTAAATGCTTCATTAGAGAA 1842
Db 613 LeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGlu 632
QY 1843 CGAATAGTCCAGTCCAGATCATGGTACTTAGACACAGTACCA 1887
Db 633 ArgValSerAlaValGlnIleMetValLeuArgGlnTrpGln 647
```

## RESULT 6

US-09-171-553B-10

; Sequence 10, Application US/09171553B

; Patent No. 6756227

; GENERAL INFORMATION:

```
; APPLICANT: GALBRAITH, DANIEL N.
; APPLICANT: HAMORTH, CHRISTINE
; APPLICANT: LEES, GILLIAN M.
; APPLICANT: SMITH, KENNETH T.
; TITLE OF INVENTION: PORCINE RETROVIRUS
; FILE REFERENCE: CFV-5.01
; CURRENT APPLICATION NUMBER: US/09/171.553B
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: PCT/GB97/01087
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: GB 9702668.6
; PRIOR FILING DATE: 1997-02-10
; PRIOR APPLICATION NUMBER: GB 9608164.1
; PRIOR FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 657
; TYPE: PRN
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Amino acid sequence
; OTHER INFORMATION: of ENV region "Raji"
US-09-171-553B-10
```

## Alignment Scores:

Pred. No.:	3,85e-233	Length:	657
Score:	2395.00	Matches:	451
Percent Similarity:	80.3%	Conservative:	75
Best Local Similarity:	68.9%	Mismatches:	95
Query Match:	69.1%	Indels:	34
DB:	2	Gaps:	10

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-09-171-553B-10 (1-657)

```
QY 1 ATGCATCCACGTTAAACCGCGCCACCTCCCGATTCGGGTGGAAGCGAAAGACTG 60
Db 1 MetHisProThrLeuSerTrpArgHisLeuProThrArgGlyGlyGluProLysArgLeu 20
QY 61 AAATCCCTTAAGCTTCGCTCCATCGCTGCTTCTTACTCTGTCAATAACCTCTCAG 120
Db 21 ArgIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrIleThrProGln 40
QY 121 ACTAATGGTATGCGCATAGGAGACAGCTCAACTCCCATAAACCTTATCTCTCACCTGG 180
Db 41 AlaSerSerLysArgLeuIleAspSerSerAsnProHisArgProLeuSerProThrTrp 60
QY 181 TTAATTACTACTCCCGCACAGGTATTAAATATCAACAACACTCAAGGGAGGCTCCTTTA 240
Db 61 LeuIleAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProArg 80
QY 241 GGAACCTGGTGGCTGATCTATAGTTTGCCTCAGATCAGTATTCTCTAGCTCAGCTCA 300
Db 81 GlyThrTrpTrpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100
QY 301 ---CCCCAGATATCTCCATGCTCAGGATTTTATGTTTCCAGGACACCAATAAT 357
Db 101 ThrProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGlu 120
QY 358 GGAACAATTCGCGAAATCCAGAGATTTCTTTTGTAAACAATGGAACTGTGTAACCTCT 417
Db 121 ---LysTyrCysGlyGlySerGlyGluSerPheCysArgArgTrpSerCysValThrSer 139
QY 418 AATGATGATATTGGAATGGCCAACTCTCAGCAGATAGGTTAGTTTCTTATGTC 477
Db 140 AsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159
QY 478 AACACCTATACAGCTCTGGACAAATTTAATTACTGACCTGGATTAGAACTGGAAGCCCC 537
Db 160 Asn-----SerGlyProGlyLysTyrLysMetLysLeuTyrLysAspLysSer--- 176
QY 538 AAGTGTCTCTCTTCAGACCTTAGATTACCTAAAAAATAGTTTTCACCTGAGAAGGAAAA 597
Db 538 AAGTGTCTCTCTTCAGACCTTAGATTACCTAAAAAATAGTTTTCACCTGAGAAGGAAAA 597
```

```
Db 177 --CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGln 195
Qy 598 GAAATATCCTAAATGGTAAATGGTATGCTTGGGGAATGGTATAT-----TATGGA 651
Db 196 GluAsnIleGlnLysTrpIleAsnGlyMetSerTrpGlyIleValPheTyrLysTyrGly 215
Qy 652 GGCTCGGTAAACAACACGAGGTCCATTCTAACTATTTCGCTCAAAATAAACACAG---CTG 708
Db 216 GlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 232
Qy 709 GAGCTCCCAATGGCTATAGACCAATACGGTCTTACGGGTCAAAGACCCCAACCCAA 768
Db 233 GluProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 252
Qy 769 GGACCA-----GGA 777
Db 253 GluProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrGln 272
Qy 778 CCATCCTCTTAACATAACTTCTGGATCAGACCCCACTGAGTCT-----ACAGC- 825
Db 273 ProProSerAsnSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 292
Qy 826 ---ACGACTAAATGGGGCAAAACTTTTACCTCATCCAGGGAGCTTTTCAAGCTCTT 882
Db 293 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 312
Qy 883 AACTCCAGACTCCAGAGGCTACTCTTCTTGTGGCTATGCTTAGCTTCGGGGCCACCT 942
Db 313 AsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProPro 332
Qy 943 TACTATGAAGAAATGGCTAGAGGGAATTCATGTGACAAAGAACATAGAGACCAA 1002
Db 333 TyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGln 352
Qy 1003 TGCATCGGGATCCCAAAATAAGCTTACCCTTACTGAGGTTTCTCGAAAGGACCTGC 1062
Db 353 CysThrTrpGlySerArgAsnLysLeuThrGluValSerGlyLysGlyThrCys 372
Qy 1063 ATAGAAAGGTTCCCGATCCCAACACACTTGTGTAAACACACTGAAGCCTTTAATCAA 1122
Db 373 IleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 392
Qy 1123 ACCTCTGAGATCAATATCTGTACTCTGTTATGACAGGTGTGGCATGTAATACTGGA 1182
Db 393 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGly 412
Qy 1183 TTAACCCCTTGTGTTCCACCTTGGTGTGTTTAAACCAACTAAAGATTTTTCATTATGTC 1242
Db 413 LeuThrProCysValSerThrSerValPheAsnGlnSerLysAspLeuCysValMetVal 432
Qy 1243 CAATTTGTTCCGGAGTGTTACTATATCCCGAAAGCAATCCTTGATGAATATGACTAC 1302
Db 433 GlnIleValProArgValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyr 452
Qy 1303 AGAATCATCCACAAACAGAGAACCCATATCTCTGACACTTGTCTGTCATGCTCGACTT 1362
Db 453 ArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeu 472
Qy 1363 GGAGTGGCAGCAGGTGTAGGAACAGGAACAGCTGCTGCTGTCAGGGACACAGAGCTA 1422
Db 473 GlyThrAlaValGlyValGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 492
Qy 1423 GAAACAGACTTAGTAACCTACATCGAATTTGAACAGAAGATCTCCAAGCCCTAGAAAAA 1482
Db 493 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlu 512
Qy 1483 TCTGTCACTAACCTGGAGAAATCCCTTAACCTTCTTATCTGAAGTAGTCTCACAGATAGA 1542
Db 513 SerValSerAsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArg 532
Qy 1543 AGAGGTTTAGATTTTATTTTCTAAAGAGGAGGATTTATGTGTAGCCTTTGAAGGAGAA 1602
Db 533 ArgGlyLeuAspLeuPheLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGlu 552
```

```
Qy 1603 TGCTGTTTTTATGTGGATCATTTCAAGGGCCATCAGAGACTCCATGAACAAGCTTTAGAGAA 1662
Db 553 CysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLys 572
Qy 1663 AGGTTGGAGAAGCCTCGAAGGGAAGAACTACTCAAGGGTGGTTTGAAGGGATGGTTC 1722
Db 573 LysLeuGluArgArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTyrPhe 592
Qy 1723 AACAGGCTCTTTGGTTGGCTACCCCTACTTCTCTCTTTAACAGGACCTTAAATAGTCTC 1782
Db 593 AsnArgSerProTyrMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu 612
Qy 1783 CTCCTGTTACTCACAGTTGGGCCATGTATTATTAAACAAGTTAAATTGCCTTTCATTAGAGAA 1842
Db 613 LeuLeuLeuLeuThrValGlyProCysLeuIleAsnArgPheValAlaPheValArgGlu 632
Qy 1843 CGAATAAGTCAGTCCAGATCATGTACTTAGACAACAGTACCAA 1887
Db 633 ArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGln 647

RESULT 7
US-09-376-781-3
; Sequence 3, Application US/09376781
; Patent No. 6261806
; GENERAL INFORMATION:
; APPLICANT: Banerjee, Papia T.
; APPLICANT: Patience, Clive
; APPLICANT: Andersson, Goran K.
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
; Patent No. 6261806
; TITLE OF INVENTION: Use
; FILE REFERENCE: 61750-267
; CURRENT APPLICATION NUMBER: US/09/376,781
; CURRENT FILING DATE: 1999-08-18
; EARLIER APPLICATION NUMBER: 60/097,015
; EARLIER FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Amino acid
; OTHER INFORMATION: sequence of a PERV env polypeptide previously
; OTHER INFORMATION: described
US-09-376-781-3

Alignment Scores:
Pred. No.: 6,81e-229 Length: 656
Score: 2353.00 Matches: 450
Percent Similarity: 79.8% Conservative: 74
Best Local Similarity: 68.5% Mismatches: 96
Query Match: 67.8% Indels: 37
DB: 2 Gaps: 10

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-09-376-781-3 (1-656)
Qy 1 ATGCATCCCACTGTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAACCGAAAGACTG 60
Db 1 MethiSProThrLeuSerSerArgHisLeuProThrArgGlyGlyGluProLysArgLeu 20
Qy 61 AAAATCCCTTAAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCAATCACTCTCAG 120
Db 21 ArgileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrIleThrProGln 40
Qy 121 ACTAATGATGCGCATAGGACAGCGCTGAACCTCCATAAACCTTATCTCTCACCTGG 180
Db 41 AlaserSerLysArgLeuIleAspSerSerAsnProHisArgProLeuSerLeuThrTrp 60
Qy 181 TTAATTAATCTACTCCGGCACAGAGTATTAAATATCAACAACACTCAAGGGAGGCTCTTTA 240
Db 181 TTAATTAATCTACTCCGGCACAGAGTATTAAATATCAACAACACTCAAGGGAGGCTCTTTA 240
```

Db 61 Leu1le1leAaspProAaspThrGlyValThrValAenSerThrArgGlyValAlaProArg 80  
Qy 241 GGAACCTGGTGGCTGATCTATACGTTGGCTCAGATCAGTATTCTTCTAGTCTGACCTCA 300  
Db 81 GlyThrTTPProGluLeuHisPheCysLeuArgLeuLeuAenProAlaValLysSer 100  
Qy 301 ---CCCCAGATATCTCATGCTCAGGATTTATGTTTGGCCAGGACCCAAATAAT 357  
Db 101 ThrProAenLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGlu 120  
Qy 358 GGAACCAATTCGGGAATCCAGAGATTTCTTTCTAAACAATGAACTGTGTAACTCT 417  
Db 121 ---LysTyrCysGlySerGlyGluSerPheCysArgArgTrpSerCysValThrSer 139  
Qy 418 AATGATGATATGGAATGCCCAACCTCTCAGCAGGATAGGTAAGTTTCTTATGTC 477  
Db 140 AenAaspGlyAaspTTPProLysTTPProLysSerLeuGlnAaspArgValLysPheSerPheVal 159  
Qy 478 AACACCTATACAGCTCTGGACAATTTAATTACCTGACCTGGATTAGAACTGGAGCCCC 537  
Db 160 Aen-----SerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAaspLysSer--- 176  
Qy 538 AAGTGCTCTCTCAGACCTAGATTACCTAAATAAGTTTCACTCAGAGAGGAACAA 597  
Db 177 ---CysSerProSerAaspLysTyrLysLysLysSerPheThr--GluArgLysThrG 195  
Qy 598 GAAATATCTTAAATGGTAAATGATGCTTGGGGAATGGTATAT-----TATGGA 651  
Db 195 LysTyrSerLysValAaspLysTyrTyrGluLeuGlyAenSer-PheLeuLeuTyrGly 214  
Qy 652 GGCTCGGTAAACCAACAGGCTCAATTTCACTATTCGCTCAAAATAAACAG--CTG 708  
Db 215 GlyGlyAla-----GlySerThrLeuThrileArgLeuArgileGluThrGlyThr 231  
Qy 709 GAGCTCCAATGGCTATAGGACCAATAAGCTCTGACGGTCAAGACCCCAACCA 768  
Db 232 GluProValAlaMetGlyProAaspLysValLeuAlaGluGlnGlyProProAlaLeu 251  
Qy 769 GGACCA-----GGA 777  
Db 252 GluProHisAenLeuProValProGlnLeuThrSerLeuArgProAaspIleThrGln 271  
Qy 778 CCATCTCTTAACATAACTTCTGGATCAGACCCCACTGAGTCT-----ACAGC----- 825  
Db 272 ProProSerAenSerThrGlyLeuIleProThrAenThrProArgAenSerProGly 291  
Qy 826 ---ACGACTAAATGGGGCAAACTTTTACCTCATCCAGGAGCTTTTCAAGCTCTT 882  
Db 292 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 311  
Qy 883 AACCTCCAGACTCCAGAGGCTACCTCTCTTGTGGCTATGCTTAGCTTCGGGCCACCT 942  
Db 312 AenSerThrAaspProAaspAlaThr-SerSerCysTrpLeuCysLeuSerSerGlyProPro 331  
Qy 943 TACTATGAAGGAATGGCTAGAGAGGGAATTCATGTGACAAAGAACATAGACCA 1002  
Db 332 TyrTyrGluGlyMetAlaLysGluArgLysPheAenValThrLysGluHisArgAenGln 351  
Qy 1003 TGCATATGGGATCCCAATAAGCTTACCTTACTGAGTTTCTGGAAGAGGACCTGC 1062  
Db 352 CysThrTrpGlySerArgAenLysLeuThrLeuThrGluValSerGlyLysGlyThrCys 371  
Qy 1063 ATAGGAAGGTTCCCCATCCCAACACCTTTGTAACCACTGAAGCTTTAATCAA 1122  
Db 372 IleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 391  
Qy 1123 ACCTCTCAGAGTCAATATCTGGTACCTGGTTATGACAGGTTGGTGGCATGTAATCTGGA 1182  
Db 392 AlaSerGluAenGlnTyrLeuValProGlyTyrAenArgTrpTrpAlaCysAenThrGly 411  
Qy 1183 TTAACCCCTGTGTTCACCTCGTTTAAACCAATAAGATTTTTCATTTATGATGTC 1242  
Db 412 LeuThrProCysValSerThrSerValPheAenGlnSerLysAaspPheCysValMetVal 431

Qy 1243 CAAATTGTTCCCGAGTGTTATTACTATCCGAAAAAGCAATCCTTGTGATGAATATGACTAC 1302  
Db 432 GlnIleValProArgValTyrTyrHisProGluGluValValLeuAaspGluTyrAaspTyr 451  
Qy 1303 AGAATCATCGACAAAAGAGAGAACCCATATCTCTGACACATGCTGTGTGATGCTCGACTT 1362  
Db 452 ArgTyrAenArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeu 471  
Qy 1363 GGAGTGGCAGCAGGTGTAGGAACAGGACACCTCCCTGGTTCAGGGACACAGCAGCTA 1422  
Db 472 GlyThrAlaValGlyValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeu 491  
Qy 1423 GAAACAGGACTTAGTAACCTACATCGAATTTTAAACAGAGATCTCCAAAGCCCTCAAAAA 1482  
Db 492 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAaspLeuArgAlaLeuLysGlu 511  
Qy 1483 TCTGTCAAGTAACCTGGAGGAATCCCTAACCTCTTATCTGAAGTAGTCTCAGAAATAGA 1542  
Db 512 SerValSerAenLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAenArg 531  
Qy 1543 AGAGGTTAGATTTATTATTCTAAAGAGGAGGATTTATGTTAGCTTCAAGAGGAGAA 1602  
Db 532 ArgGlyLeuAaspLeuLeuPheLeuArgGluGlyLeuCysAlaAlaLeuLysGluGlu 551  
Qy 1603 TGCTGTTTATGTGGATCATTCAAGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAA 1662  
Db 552 CysCysPheTyrValAaspHisSerGlyAlaIleArgAaspSerMetAenLysLeuArgLys 571  
Qy 1663 AGTTGGAGAGCTCGAAGGAAAGGAAACTACTCAAGGGTGGTTTGAAGGATGGTTC 1722  
Db 572 LysLeuGluArgArgArgGluArgGluAlaAaspGlnGlyTrpPheGluGlyTrpPhe 591  
Qy 1723 AACAGGCTCTTGGTTGGCTTACCTACTTCTCTTAAACAGACCCCTTAATAGTCTC 1782  
Db 592 AenArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu 611  
Qy 1783 CTCCTGTACTCAGAGTTGGGCCATGTATTATTAAACAAGTTAATTCCTTCAATAGAGAA 1842  
Db 612 LeuLeuLeuLeuThrValGlyProCysLeuIleAenArgPheValAlaPheValArgGlu 631  
Qy 1843 CGAATAAGTCAGTCAGATCATGTTACTTAGACAAACAGTACCAA 1887  
Db 632 ArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGln 646

## RESULT 8

US-09-171-553B-6  
; Sequence 6, Application US/09171553B  
; Patent No. 6756227  
; GENERAL INFORMATION:  
; APPLICANT: GALBRAITH, DANIEL N.  
; APPLICANT: HAWORTH, CHRISTINE  
; APPLICANT: LEES, GILLIAN M.  
; APPLICANT: SMITH, KENNETH T.  
; TITLE OF INVENTION: PORCINE RETROVIRUS  
; FILE REFERENCE: CFV-5.01  
; CURRENT APPLICATION NUMBER: US/09/171.553B  
; CURRENT FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: PCT/GB97/01087  
; PRIOR FILING DATE: 1997-04-18  
; PRIOR APPLICATION NUMBER: GB 9702668.6  
; PRIOR FILING DATE: 1997-02-10  
; PRIOR APPLICATION NUMBER: GB 9608164.1  
; PRIOR FILING DATE: 1996-04-19  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 656  
; TYPE: PRT  
; ORGANISM: Porcine retrovirus

Alignment Scores:

Pred. No.: 6,81e-229 Length: 656  
Score: 2353.00 Matches: 450  
Percent Similarity: 79.8% Conservative: 74  
Best Local Similarity: 68.5% Mismatches: 96  
Query Match: 67.8% Indels: 37  
DB: 2 Gaps: 10

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-09-171-553B-6 (1-656)

Qy 1 ATGCATCCAGCTTAAACCGCGCCACCTCCGATTCGGGTGGAAAGCCGAAAGACTG 60  
Db 1 MethisProThrLeuSerArgHisLeuProThrArgGlyGlyGluProLysArgLeu 20  
Qy 61 AAATATCCCTTAAGCTTCGCTCCATCGCGGTTCCTTACTCTCAATAACCTCTCAG 120  
Db 21 ArgileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuThrIleProGln 40  
Qy 121 ACTAATGCTATCGCATAGGAGACAGCTGAACTCCCATAAACCTTATCTCTACCTGG 180  
Db 41 AlaSerSerLysArgLeuLeuAspSerSerAsnProHisArgProLeuSerLeuThrTrp 60  
Qy 181 TTAATTAAGTATCCGCGCACAGGTATTAATATCAACAACACTCAAGGAGGCTCTTTA 240  
Db 61 LeuIleLeuAspProAspThrGlyValThrValAsnSerThrArgGlyValAlaProArg 80  
Qy 241 GGAACCTGGTGGCTGATACGTATTCGCTCAGATCAGTATTTCTAGTCTGACCTCA 300  
Db 81 GlyThrTrpTrpProGluLeuHisPheCysLeuArgLeuIleAsnProAlaValLysSer 100  
Qy 301 ---CCCCAGATATCCTCATGCTCAGCGATTTATGTTTCCAGGACACCAAAATAAT 357  
Db 101 ThrProProAsnLeuValArgSerTyrGlyPheTyrCysCysProGlyThrGluLysGlu 120  
Qy 358 GGAAACAATTCGGGAAATCCAGAGATTCTTTTCTAAACAATGAATGTTGTAACCTCT 417  
Db 121 ---LysTyrCysGlySerGlyGluSerPheCysArgArgTrpSerCysValThrSer 139  
Qy 418 AATGATGATATGGAATGGCCACCTCTCAGCAGGATAGGTAAAGTTTCTTATGTC 477  
Db 140 AsnAspGlyAspTrpLysTrpProIleSerLeuGlnAspArgValLysPheSerPheVal 159  
Qy 478 AACACCTATACAGCTCTGGACAAATTAATACCTGACCTGATAGAACTGGAAAGCCCC 537  
Db 160 Asn-----SerGlyProGlyLysTyrLysMetMetLysLeuTyrLysAspLysSer--- 176  
Qy 538 AAGTCTCTCTCAGACTAGATTACTTAAATAAGTTTCACTGAGAAAGGAAACAA 597  
Db 177 ---CysSerProSerAspLeuAspTyrLeuLysIleSerPheThr--GluArgLysThrG 195  
Qy 598 GAAATATCCTAAATGGTAAATGGTATGCTTGGGGAATGGTATAT-----TATGGA 651  
Db 195 LysLysTyrSerLysValAspLysTrpTyrGluLeuGlyAsnSer--PheLeuLeuTyrGly 214  
Qy 652 GCCTCGGTAAACAACACAGGCTCCATTCTAATATTCCGCTCMAAATAAACCCAG---CTG 708  
Db 215 GlyGlyAla-----GlySerThrLeuThrIleArgLeuArgIleGluThrGlyThr 231  
Qy 709 GAGCTCCCAATGGCTATAGGACCAATAACGGTCTTTGACGGGTCAAAGACCCCAACCAA 768  
Db 232 GluProProValAlaMetGlyProAspLysValLeuAlaGluGlnGlyProProAlaLeu 251  
Qy 769 GGACCA-----GGA 777  
Db 252 GluProProHisAsnLeuProValProGlnLeuThrSerLeuArgProAspIleThrGln 271  
Qy 778 CCATCCTCTAAACATAACTTCTGGATCAGACCCCACTGAGTCT-----AACAGC----- 825  
Db 272 ProProSerAsnSerThrThrGlyLeuIleProThrAsnThrProArgAsnSerProGly 291  
Qy 826 ---ACGACTAAATCGGCGCAAACTTTTACCTCATCCAGGAGCTTTTCAAGCTCTT 882  
Db 292 ValProValLysThrGlyGlnArgLeuPheSerLeuIleGlnGlyAlaPheGlnAlaIle 311

## RESULT 9

US-09-315-127-5

; Sequence 5, Application US/09315127

; Patent No. 6448390

; GENERAL INFORMATION:

Qy 883 AACTCCAGACTCCAGAGGCTACCTCTTCTTGTGGCTATCTTAGCTTCGGGCCACCT 942  
Db 312 AsnSerThrAspProAspAlaThrSerSerCysTrpLeuCysLeuSerSerGlyProPro 331  
Qy 943 TACTATGAAGGAATGGCTAGAGGGAATTCATATGTGACAAAAGAACATAGAGACCAA 1002  
Db 332 TyrTyrGluGlyMetAlaLysGluArgLysPheAsnValThrLysGluHisArgAsnGln 351  
Qy 1003 TGCATGCGGATCCCAATAAAGCTTACCTTACTGAGGTTCCTGGAAGAGGACCCGTC 1062  
Db 352 CysThrTrpGlySerArgAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCys 371  
Qy 1063 ATAGAAAGGTTCCTCCATCCCAACACCTTTGTAAACACACTGAAGCCTTTAATCAA 1122  
Db 372 IleGlyLysAlaProProSerHisGlnHisLeuCysTyrSerThrValValTyrGluGln 391  
Qy 1123 ACCTCTGAGATCAATATCTGGTACTCTGTTATGACAGGTGGTGGCATGTAATCTGGA 1182  
Db 392 AlaSerGluAsnGlnTyrLeuValProGlyTyrAsnArgTrpTrpAlaCysAsnThrGly 411  
Qy 1183 TTAACCCCTTGTGTTCCACCTTGGTTTTAACCAACTAAAGATTTTTCATTTATGCTC 1242  
Db 412 LeuThrProCysValSerThrSerValPheAsnGlnSerLysAspPheCysValMetVal 431  
Qy 1243 CAAATTTGTTCCCGAGTGTATTACTATCCCGAAAAAGCAATCCTTTGATGATATGACTAC 1302  
Db 432 GlnIleValProArgValTyrTyrHisProGluGluValValLeuAspGluTyrAspTyr 451  
Qy 1303 AGAATCATCGACAAAAAGAGAGAACCCATATCTCTGACACTTGTCTGTGATGCTCGGACT 1362  
Db 452 ArgTyrAsnArgProLysArgGluProValSerLeuThrLeuAlaValMetLeuGlyLeu 471  
Qy 1363 GGAGTGGCAGCAGCTGTAGGAACAGGACAGCTCCCTGCTCAGGGACACACAGCAGTA 1422  
Db 472 GlyThrAlaValGlyValGlyThrGlyThrAlaLeuIleThrGlyProGlnGlnLeu 491  
Qy 1423 GAAACAGACTTAGTAACTACATCGAATTTCTAAACAGAAAGATCTCCAAAGCCTTAGAAAA 1482  
Db 492 GluLysGlyLeuGlyGluLeuHisAlaAlaMetThrGluAspLeuArgAlaLeuLysGlu 511  
Qy 1483 TCTGTGAGTAACCTGGAGGAATCCCTAACCTCTTATCTGAAGTAGTCTCTCAGAAATAGA 1542  
Db 512 SerValSerAsnLeuGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArg 531  
Qy 1543 AGAGGTTAGATTATTTATTTCTAAAGAGAGGATTTATGTTAGCTTCGAAGGAGAA 1602  
Db 532 ArgGlyLeuAspLeuLeuArgGluGlyGlyLeuCysAlaAlaLeuLysGluGlu 551  
Qy 1603 TGCTGTTTTTATGTGGATCATTCAGGGGCCATCAGAGACTCCATGAAACAAGCTTAGAGAA 1662  
Db 552 CysCysPheTyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgLys 571  
Qy 1663 AGGTTGGAGAGCGTCGAAGGGAAGAAACTACTCAAGGGTGGTTTGGAGGATGGTTC 1722  
Db 572 LysLeuGluArgArgArgGluArgGluAlaAspGlnGlyTyrPheGluGlyTrpPhe 591  
Qy 1723 AACAGTCTCTTGTGGTCTACCTACTTTCTGCTTTAAACAGGACCCCTTAATAGCTCTC 1782  
Db 592 AsnArgSerProTrpMetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeu 611  
Qy 1783 CTCCTGTTTACTCACAGTTGGGCCATGTATTATTAAAGATTAATTCCTTTCATTAGAGAA 1842  
Db 612 LeuLeuLeuLeuThrValGlyProCysLeuLeuAsnArgPheValAlaPheValArgGlu 631  
Qy 1843 CGAATAAGTCAGTCCAGATCATGTTAGTACTTATAGAACCAACATACCAA 1887  
Db 632 ArgValSerAlaValGlnIleMetValLeuArgGlnGlnTyrGln 646

```

; APPLICANT: The University of Tennessee, c/o Richard Cox
; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
; FILE REFERENCE: Liposome Vectors and Use in Gene and Drug Therapy
; CURRENT APPLICATION NUMBER: US/09/315,127
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-315-127-5

Alignment Scores:
Pred. No.: 1,55e-141 Length: 667
Score: 1489.50 Matches: 314
Percent Similarity: 58.0% Conservative: 90
Best Local Similarity: 45.1% Mismatches: 171
Query Match: 42.9% Indels: 121
DB: 2 Gaps: 20

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-09-315-127-5 (1-667)

QY 10 ACCTTAACCCGGCCACCTC-----CCGATTGGGGTGGAAAGCCGAAAGACTGAAA 63
DB 11 ThrSerAsnLeuHisHisLeuArgHisGlnMetSerProGlySerTrpLysArgLeu--- 29
QY 64 ATCCCTTAAGCTTCGCGCTCCATCGCGTGGTTCTTACTCTGTCGTAACCTCTCAGACT 123
DB 30 -----llelleuLeuSerCysValPheGlyGly 39
QY 124 AATGGTATGCGCATAGGAGACAGCTGAACTCCCATAAACCTTTATCTCTCACCTGGTTA 183
DB 40 GlyGlyThrSerLeu---GlnAsnLysAsnProHisGlnProMetThrLeuThrTrpGln 58
QY 184 ATTAAGTCTCCGGACAGGTATTAATATCAACAACTCAAGGGGAGGCTCTTTAGGA 243
DB 59 ValLeu---SerGlnThrGlyAspValValTrpAspThrLysAlaValGlnProProTrp 77
QY 244 ACCTGGTGGCTGATCTATAGTTGGCTCAGATCAGTATTCTCTAGCTGACCTCA--- 300
DB 78 ThrTrpProThrLeuLysProAspValCysAlaLeuAlaLaSerLeuGluSerTrp 97
QY 301 -----CCCCAGATATC----- 312
DB 98 AspileProGlyThrAspValSerSerLysArgValArgProProAspSerAspTyr 117
QY 312 ----- 312
DB 118 ThrAlaAlaLysGlnIleThrTrpGlyAlaIleGlyCysSerTyrProArgAlaArg 137
QY 313 -----CTCCATGCTCAGGATTTTATGTTGGCCAGGACCACCA-----AATAAT 357
DB 138 ThrArgMetAlaSerSerThrPheTyrValCysProArgAspGlyArgThrLeuSerGlu 157
QY 358 GGAAGAACATGCGGAAATCCACAGATTTCTTTTGTAAACAATGGAATGTGTAACCTCT 417
DB 158 AlaArgArgCysGlyGlyLeuGluSerLeuTyrCysLysGluTrpAspCysGluThrThr 177
QY 418 AATGATGGATATTGG-----AAATGCCCAACC 444
DB 178 GlyThrGlyTyrTrpLeuSerLysSerSerLysAspLeuIleThrValLysTrpAspGln 197
QY 445 TCTCAGCAGGATAGGTAAAGTTTCTTATGTCAACCTATACCACTCTGGACATTT 504
DB 198 AsnSerGluTrpThrGlnLysPheGln-----GlnCys 208
QY 505 AATTACCTGACCTGGATTAGAACTGGAAGCCCAAGTGTCTCTCTTCAGACCTAGATTAC 564
DB 209 HisGlnThrGlyTrp-----CysAsnPro----- 216
QY 565 CTAAGAAATAAGTTTCACTGAGAAAGGAAAAACAAGAAATATCTCTAAATGGGTAAATG 624

```

```

||||| 217 LeuLysIleAspPheThrAspLysGlyLysLeuSerLys-----AspTrpIleThrGly 234
QY 625 ATGCTTGGGGAATGGTATATTATGGAGGCTCGGTTAAACAACAGGCTCCATTTCTACT 684
DB 235 LysThrTrpGlyLeuArgPheTyr---ValSerGly---HisProGlyValGlnPheThr 252
QY 685 ATTCGCTCAAAATAAAACACGAGCTCGAGCTCCAATGGCTATAGACCAAAATACGGCTCTG 744
DB 253 IleArgLeuLysIleThrAsnMet---ProAlaValAlaValGlyProAspLeuValLeu 271
QY 745 ACGGTCAA-----AGACCCCCAACCCAA 768
DB 272 ValGluGlnGlyProProArgThrSerLeuAlaLeuProProLeuProProArgGlu 291
QY 769 GGACGAGGACCATCC-----TCTAAC-----ATAACTCTGGATCAGAC 807
DB 292 AlaProProSerLeuProAspSerAsnSerThrAlaLeuAlaThrSerAlaGlnThr 311
QY 808 CCCACTGAGTCTAACAGCAGCAGCTAAATG-----GGGGCAAAA 846
DB 312 ProThrValArgLysThrIleValThrLeuAsnThrProProThrThrGlyAspArg 331
QY 847 CTTTTAGCCTATCCAGGAGCTTTTCAAGCTCTTAAGCTTAACTCCAGACTCCAGAGGCTACC 906
DB 332 LeuPheAspLeuValGlnGlyAlaPheLeuThrLeuAsnAlaThrAsnProGlyAlaThr 351
QY 907 TCTTCTTGGTGTATGCTTGTAGCTTCGGGCCACCTTACTATGAAGNAATGGCTAGAAGA 966
DB 352 GluSerCysTrpLeuCysLeuAlaMetGlyProProTyrTyrGluAlaIleAlaSerSer 371
QY 967 GGGAAATTCATATGTACAAAAGAACATAGAGACCAATGCATCGGATGCCAAATAAAG 1026
DB 372 GlyGluValAlaLysSerThrAsp---LeuAspArgCysArgTrpGlyThrGlnGlyLys 390
QY 1027 CTTACCTTACTGAGGTTTCTGGAAGAGCCCTGCATAGAGAAAGTTGCCCAATCCCAAC 1086
DB 391 LeuThrLeuThrGluValSerGlyHisGlyLeuCysIleGlyLysValProPheThrHis 410
QY 1087 CACACCTTGTAAACACACTGAAGCCTTTAATCAACCTCTGAGAGTCAATATCTGGTA 1146
DB 411 GlnHisLeuCysAsnGlnThrLeuSerIleAsnSerSerGlyAspHisGlnTyrLeuLeu 430
QY 1147 CCTGGTTATCAGAGGTGGTGGCATGTAATCTGGAATTAACCCCTGTGTTTCCACCTTG 1206
DB 431 ProSerAsnHisSerTrpTrpAlaCysSerThrGlyLeuThrProCysLeuSerThrSer 450
QY 1207 GTTTTTAACCAACTAAAGATTTTGTTCATTATGGTCCAAATTTGTTCCCGAGTGATTTAC 1266
DB 451 ValPheAsnGlnThrArgAspPheCysIleGlnValGlnLeuIleProArgIleTyr 470
QY 1267 TATCCCGAAAAGCAATCCTTGATGATATGATGACTACAGAAATCATCCACAAAGAGAGAA 1326
DB 471 TyrProGluGluValLeuLeuGlnAlaTyrAspAsnSerHisProArgThrLysArgGlu 490
QY 1327 CCATATCTCTGACACTTGTGATGCTCGAGTCTCGAGTGGCAGCAGGTGTAGGAAACA 1386
DB 491 AlaValSerLeuThrLeuAlaValLeuLeuGlyLeuGlyIleThrAlaGlyIleGlyThr 510
QY 1387 GGAACAGCTGCTTGTGTCAGGACCAACAGCAGCTAGAGAAACAGAGCTTAGTAACCTACAT 1446
DB 511 GlySerThrAlaLeuIleLysGlyProIleAspLeuGlnGlnGlyLeuThrSerLeuGln 530
QY 1447 CGAATTGTAAACAGAAATCTCCAAAGCCCTAGAAAAATCTGCTAGTAACCTGGAGGATCC 1506
DB 531 IleAlaIleAspAlaAspLeuArgAlaLeuGlnAspSerValSerLysLeuGluAspSer 550
QY 1507 CTAACCTCTTATCTGAAGTAGTCTACAGAAAGAGAGGTTAGATTTATTATTCTA 1566
DB 551 LeuThrSerLeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeuPheLeu 570
QY 1567 AAAGAGAGGAGGATTATGTGTAGCTTGAAGGAGAAATGCTGTTTTTATGTCATTCATTC 1626

```



```
Db 571 LysGluGlyGlyLeuCybAlaLeuLysGluGlyCysPheThrIleAspHisSer 590
Qy 1627 GGGCCATCAGACTCCATGAACAGCTTAGAGAAAGTTGGAGAACGCTCGAAGGAA 1686
Db 591 GlyAlaValArgAspSerMetLysLysLeuLysGluLysLeuAspLysArgGlnLeuGlu 610
Qy 1687 AAGGAAACTACTCAAGGGTGGTTGGAGGATGGTTCAACAGGTCTCTTTGGTGGCTACC 1746
Db 611 ArgGlnLysSerGlnAsnTrpArgGluGlyTrpPheAsnAsnSerProTrpPheThr 630
Qy 1747 CTACTTTCTGCTTTAAACAGGACCCCTTAATAGTCCTCTCTCTCTACTCAGAGTTGGGCA 1806
Db 631 LeuLeuSerThrIleAlaGlyProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuGlyPro 650
Qy 1807 TGTATTATTAACAGTTAATTCCTTCAATTTAGAGAACGAATAAGTGCA 1854
Db 651 CysIleIleAsnLysLeuValGlnPheIleAsnAspArgIleSerAla 666

RESULT 10
US-09-315-127-6
; Sequence 6, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
; APPLICANT: The University of Tennessee, c/o Richard Cox
; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
; TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
; FILE REFERENCE: 44137-5023, U. of Tennessee
; CURRENT APPLICATION NUMBER: US/09/315,127
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQ ID NO. 4,
; OTHER INFORMATION: envelope protein produced by retroviral vector of
; OTHER INFORMATION: seq. id no. 3
US-09-315-127-6

Alignment Scores:
Pred. No.: 1,55e-141 Length: 667
Score: 1489.50 Matches: 314
Percent Similarity: 58.0% Conservative: 90
Best Local Similarity: 45.1% Mismatches: 171
Query Match: 42.9% Indels: 121
DB: 2 Gaps: 20

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-09-315-127-6 (1-667)
Qy 10 ACGTTAAACCGCGCCACCTC-----CCGATTCCGGGTGGAAAGCCGAAAGACTGAAA 63
Db 11 ThrSerAsnLeuHisLeuArgHisGlnMetSerProGlySerTrpLysArgLeu--- 29
Qy 64 ATCCCTTAAGCTTCGCCTCCATCCGCGGTTCCTTACTCTGTCTAATAACCTCTCAGACT 123
Db 30 -----fLeileLeuLeuSerCysValPheGlyGly 39
Qy 124 AATGGTATCGGCATAGGACAGACCTGAACCTCCCATAAACCTTATCTCCTACCTGGTTA 183
Db 40 GlyGlyThrSerLeu---GlnAsnLysAsnProHisGlnProMetThrLeuThrTrpGln 58
Qy 184 ATTACTGACTCCGGCAGCAGGTATTATATCAACAACTCAAGGGGAGGCTCTTTAGGA 243
Db 59 ValLeu---SerGlnThrGlyAspValValTrpAspThrLysAlaValGlnProProTrp 77
Qy 244 ACCTGGTGGCTGATCTATACCTTTGGCTCAGATCAGTTATTCTTAGTCTGACCTCA--- 300
Db 78 ThrTrpTrpProThrLeuLysProAspValCysAlaLeuAlaAspSerLeuGluSerTrp 97
Qy 301 -----CCCCCAGATATC----- 312
```

```
Db 98 AspIleProGlyThrAspValSerSerLysArgValArgProProAspSerAspTyr 117
Qy 312 ----- 312
Db 118 ThrAlaAlaTyrLysGlnIleThrTrpGlyAlaIleGlyCysSerTyrProArgAlaArg 137
Qy 313 -----CTCCATGCTCACGGATTTTATGTTTTCGCCAGGACCACCA-----AATAAT 357
Db 138 ThrArgMetAlaSerSerThrPheTyrValCysProArgAspGlyArgThrLeuSerGlu 157
Qy 358 GGAACCATTCGCGAATCCAGAGATTCCTTTTGAACAATGGAACCTGTGTAACTCT 417
Db 158 AlaArgArgCysGlyGlyLeuGluSerLeuTyrCysLysGluTrpAspCysGluThrThr 177
Qy 418 AATGATGATATGG-----AATGGCCCAACC 444
Db 178 GlyThrGlyTyrTrpLeuSerLysSerLysAspLeuIleThrValLysTrpAspGln 197
Qy 445 TCTCAGCAGGATAGGGTAAGTTTCTTATGTCACACCTATACCAGCTCTGGCAATTT 504
Db 198 AsnSerGluTrpThrGlnLysPheGln-----GlnCys 208
Qy 505 AATTACCTGACCTGGATTAGAACTGGAAGCCCCCAAGTCTCTCTCTCAGACCTAGATTAC 564
Db 209 HisGlnThrGlyTrp-----CysAsnPro----- 216
Qy 565 CTAAAAATAGTTTCACCTAGAGAAAGGAAACAGAAATATCTCTAAATGGTAAATGGT 624
Db 217 LeuLysIleAspPheThrAspLysGlyLysLeuSerLys-----AspTrpIleThrGly 234
Qy 625 ATGCTCTGGGAATGGTATATTATGGAGGCTCGGAGTAAACCAACACCGGCTCTCACT 684
Db 235 LysThrTrpGlyLeuArgPheTyr---ValSerGly---HisProGlyValGlnPheThr 252
Qy 685 ATTCGCTCAAAATAAACACGAGCTGAGCTCCAAATGGCTATAGACCAAAATACGGTCTTG 744
Db 253 IleArgLeuLysIleThrAsnMet---ProAlaValAlaValGlyProAspLeuValLeu 271
Qy 745 ACGGTCNA-----AGNCCCCCAACCCAA 768
Db 272 ValGluGlnGlyProProArgThrSerLeuAlaLeuProProLeuProProArgGlu 291
Qy 769 GGACCAAGGACCATCC-----TCTAAC-----ATAACTTCTGGATCAGAC 807
Db 292 AlaProProSerLeuProAspSerAsnSerThrAlaLeuAlaThrSerAlaGlnThr 311
Qy 808 CCCACTGAGTCTAACAGCAGCAGCTAAATG-----GGGCAAAA 846
Db 312 ProThrValArgLysThrIleValThrLeuAsnThrProProThrThrGlyAspArg 331
Qy 847 CTTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCAGACTCCAGAGGCTACC 906
Db 332 LeuPheAspLeuValGlnGlyAlaPheLeuThrLeuAsnAlaThrAsnProGlyAlaThr 351
Qy 907 TCTTCTGCTGCTATGCTTAGCTTCGGGCCACCTTACTATGAAGGAATGGCTAGAAGA 966
Db 352 GluSerCysTrpLeuCysLeuAlaMetGlyProProTyrTyrGluAlaIleAlaSerSer 371
Qy 967 GGGAAATCAATGTGACAAAAGAACATAGACCAATGACATGGGGATCCCAAAATAG 1026
Db 372 GlyGluValAlaTyrSerThrAsp---LeuAspArgCysArgTrpGlyThrGlnGlyLys 390
Qy 1027 CTTACCTTACTGAGGTTTCTGGAAAGGACCTGATAGGAAGGTTTCCCCCATCCAC 1086
Db 391 LeuThrLeuThrGluValSerGlyHisGlyLeuCysIleGlyLysValProPheThrHis 410
Qy 1087 CAACACCTTTCTAACCACTGAGCCCTTTTAACTCAACCTCTCTGAGAGTCAATATCTGTA 1146
Db 411 GlnHisLeuCysAsnGlnThrLeuSerIleAsnSerSerGlyAspHisGlnThrLeuLeu 430
Qy 1147 CCTGTTATGACAGGTGGTGGCATGTAATACTGGATTAAACCTTGTGTTTCCACCTTG 1206
Db 431 ProSerAsnHisSerTrpTrpAlaCysSerThrGlyLeuThrProCysLeuSerThrSer 450
```

```
QY 1207 GTTTTAAACCAAACTAAGATTGTTGCAATTATGGTCCAAATGTTCCCGAGTGATTAC 1265
Db 451 ValPheAsnGlnThrArgAspPheCysIleGlnValGlnLeuIleProArgIleTyr 470
QY 1267 TATCCCGAAAAGCAATCCTTGATGAATATGATACAGAAATCATCGACAAAGAGAA 1326
Db 471 TyrProGluGluValLeuLeuGlnAlaTyrAspAsnSerHisProArgThrLysArgGlu 490
QY 1327 CCCATATCTGACACTTGCTGTGATGTCGCGACTTGGAGTGGCAGCAGGTGTAGGAACA 1386
Db 491 AlaValSerLeuThrLeuAlaValLeuLeuGlyLeuGlyIleThrAlaGlyIleGlyThr 510
QY 1387 GGAACAGCTGCTGCTGTCACGGGACCAACAGCAGCTAGAAACAGGACTTAGTAACCTACAT 1446
Db 511 GlySerThrAlaLeuIleLysGlyProIleAspLeuGlnGlnGlyLeuThrSerLeuGln 530
QY 1447 CGAATTGTAACAGAGATCTCCAGCCCTAGAAAATCTGTCAAGTAACCTGAGGATCC 1506
Db 531 IleAlaIleAspAlaAspLeuArgAlaLeuGlnAspSerValSerLysLeuGluAspSer 550
QY 1507 CTAACCTCTTATCTGAAGTAGTCTCAGAGATAGAGAGGCTTAGATTATTATTCTTA 1566
Db 551 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu 570
QY 1567 AAGAAGGAGGATTATGTAGCTTGAAGGAGGAATCTGTTTTATGTGGATCATTTCA 1626
Db 571 LysGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysPheTyrIleAspHisSer 590
QY 1627 GGGGCCATCAGAGACTCCATCAACAGCTTAGAAGAGGTTGGAGAGCGTCGAGGGAA 1686
Db 591 GlyAlaValArgSerMetLysLysLeuLysGluLysLeuAspLysArgGlnLeuGlu 610
QY 1687 AAGAAACTACTCAAGGTGCTTGGAGGATGTTCAACAGTCTCTTTGGTGGTACC 1746
Db 611 ArgGlnLysSerGlnAsnTyrTyrGluGlyTyrPheAsnAsnSerProTyrPheThrThr 630
QY 1747 CTACTTCTGCTTTAAACAGGACCTTAATAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1806
Db 631 LeuLeuSerThrIleAlaGlyProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuGlyPro 650
QY 1807 TGTATTATTAAACAGTTAATGCTTCATTTAGAGAACGAATAAGTCCA 1854
Db 651 CysIleIleAsnLysLeuValGlnPheIleAsnAspArgIleSerAla 666

RESULT 11
US-09-070-630-13
; Sequence 13, Application US/09070630
; Patent No. 6750206
; GENERAL INFORMATION:
; APPLICANT: Russell, Stephen James
; APPLICANT: Morling, Frances Joanne
; APPLICANT: Fielding, Adele Kay
; APPLICANT: Cosset, Francois-Loic
; APPLICANT: Cattaneo, Roberto
; TITLE OF INVENTION: Compositions and Methods for Elimination
; TITLE OF INVENTION: of Unwanted Cells
; FILE REFERENCE: 07039-415001
; CURRENT APPLICATION NUMBER: US/09/070,630
; CURRENT FILING DATE: 1998-04-30
; PRIOR APPLICATION NUMBER: PCT/GB98/00710
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: US 60/045,164
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: GB 9705007.4
; PRIOR FILING DATE: 1997-03-11
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 667
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: Plasmid sequence
US-09-070-630-13
Alignment Scores:
Pred. No.: 1,55e-141 Length: 667
Score: 1489.50 Matches: 314
Percent Similarity: 58.0% Conservatives: 90
Best Local Similarity: 45.1% Mismatches: 171
Query Match: 42.9% Indels: 121
DB: 2 Gaps: 20
US-10-723-552-3_COPY_5620_7533 (1-1914) x US-09-070-630-13 (1-667)
QY 10 ACGTTTAAACCGCGCCACCTC-----CCGATTCCGGGTGGAAGCCGAAAAGACTGAAA 63
Db 11 ThrSerAsnLeuHisHisLeuArgHisGlnMetSerProGlySerTrpLysArgLeu--- 29
QY 64 ATCCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCAATAACCTCTCAGACT 123
Db 30 -----IleIleLeuLeuSerCysValPheGlyGly 39
QY 124 AATGTTATCGCATAGGAGACAGCCTGAACTCCCAATAAACCTTATCTCTCACCCTGGTTA 183
Db 40 GlyGlyThrSerLeu---GlnAsnLysAsnProHisGlnProMetThrLeuThrTrpGln 58
QY 184 ATTACTGACTCCCGCACAGGTATTAAATCAACAACACTCAAGGGGAGGCTCTCTTTAGGA 243
Db 59 ValLeu---SerGlnThrGlyAspValValTrpAspThrLysAlaValGlnProProTrp 77
QY 244 ACCTGGTGGCTGATCTATACGTTGGCTCAGATCAGTTATTCCTAGTCTGACCTCA--- 300
Db 78 ThrTrpTrpProThrLeuLysProAspValCysAlaLeuAlaLeuSerLeuGluSerTrp 97
QY 301 -----CCCCAGATATC----- 312
Db 98 AspIleProGlyThrAspValSerSerLysArgValArgProProAspSerAspTyr 117
QY 312 ----- 312
Db 118 ThrAlaAlaTyrLysGlnIleThrTrpGlyAlaIleGlyCysSerTyrProArgAlaArg 137
QY 313 -----CTCATGCTCAGGATTTTATGTTGGCCAGGACCAACA-----AATAAT 357
Db 138 ThrArgMetAlaSerSerThrPheTyrValCysProArgAspGlyArgThrLeuSerGlu 157
QY 358 GGAATAACATTCCGGAATCCAGAGATTTCTTTGTAAACAATGGAAGTGTGTAACCTCT 417
Db 158 AlaArgArgCysGlyGlyLeuGluSerLeuTyrCysLysGluTrpAspCysGluThrThr 177
QY 418 AATGATGATATTGG-----AAATGGCCAACC 444
Db 178 GlyThrGlyTyrTrpLeuSerLysSerSerLysAspLeuIleThrValLysTrpAspGln 197
QY 445 TCTCAGCAGGATAGGTAAAGTTTCTTATGTCAACACCTATACCACTCTGGCAATTT 504
Db 198 AsnSerGluTrpThrGlnLysPheGln-----GlnCys 208
QY 505 AATTACCTGACCTGGATTAGAACTGGAGCCCGCCAGTGTCTCTCTCAGACCTAGATTAC 564
Db 209 HisGlnThrGlyTrp-----CysAsnPro----- 216
QY 565 CTAATAATAAGTTTCTACTGAGAAAGAAAACAAGAAATATCTTAAATGGTAAATGGT 624
Db 217 LeuLysIleAspPheThrAspLysGlyLysLeuSerLys-----AspTrpIleThrGly 234
QY 625 ATGCTTGGGGAATGGTATATTATGGAGGCTCGGGTAAACAACAGGCTCCATTCTAACT 684
Db 235 LysThrTrpGlyLeuArgPheTyr---ValSerGly---HisProGlyValGlnPheThr 252
QY 685 ATTGCGCTCAAAATAAACACGCTGGAGCTCCCAATGGCTATAGGACCAATAACGCTTTG 744
Db 253 IleArgLeuLysIleThrAsnMet---ProAlaValAlaValGlyProAspLeuValLeu 271
```

Qy	745	ACGGGTCAA	-----AGACCCCAACCCAA	768
Db	272	ValGluGlnGlyProProArgThrSerLeuAlaLeuProProLeuProProArgGlu		291
Qy	769	GGACCAAGGACATCC	-----TCTAAC	807
Db	292	AlaProProSerLeuProProSerAsnSerThrAlaLeuAlaThrSerAlaGlnThr		311
Qy	808	CCCACTCAGTCTAACACACGACGACTAAATG	-----GGGGCAAA	846
Db	312	ProThrValArgIysThrIleValThrLeuAsnThrProProProThrThrGlyAspArg		331
Qy	847	CTTTTTAGCCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCAGACTCCAGAGGCTACC		906
Db	332	LeuPheAspLeuValGlnGlyAlaPheLeuThrLeuAsnAlaThrAsnProGlyAlaThr		351
Qy	907	TCCTCTGTGTGGCTATGCTTAGCTTCGGGGCCACCTTACTATGAGGAATGGCTAGNAGA		966
Db	352	GluSerCysTrpLeuCysLeuAlaMetGlyProProTyrTyrGluAlaIleAlaSerSer		371
Qy	967	GGGAAATTCATGTGCAAAAGAACAATAGACCAATGCACATGGGGATCCCAAAATAAG		1026
Db	372	GlyGluValAlaTyrSerThrAsp--LeuAspArgCysArgTrpGlyThrGlnGlyLys		390
Qy	1027	CTTACCCCTTACTGAGGTTCTGGAAAGCGCACTGCATAGAGAAAGTTTCCCCATCCAC		1086
Db	391	LeuThrLeuThrGluValSerGlyHisGlyLeuCysIleGlyLysValProPheThrHis		410
Qy	1087	CAACACCTTTGTAAACCACTGAAGCCTTTAATCAAACTCTGAGAGCTCAATATCTGGTA		1146
Db	411	GlnHisLeuCysAsnGlnThrLeuSerIleAsnSerSerGlyAspHisGlnTyrLeuLeu		430
Qy	1147	CCTGGTTTACAGCGTGTGGCATGTAATCTGGATTAACTGATTAACCCCTGTGTTTCCACCTG		1206
Db	431	ProSerAsnHisSerTrpTrpAlaCysSerThrGlyLeuThrProCysLeuSerThrSer		450
Qy	1207	GTTTTTAACCAACTAAAGATTTTGTGATTTATGTCCTCAAAATGTTCCTCCAGTGTATTAC		1266
Db	451	ValPheAsnGlnThrArgAspPheCysIleGlnValGlnLeuIleProArgIleTyrTyr		470
Qy	1267	TATCCCGAAAGAACATTCCTTGATGATATGACTACAGAAATCATCGACAAAGAGAGAA		1326
Db	471	TyrProGluGluValLeuLeuLeuGlnAlaTyrAspAsnSerHisProArgThrLysArgGlu		490
Qy	1327	CCCATATCTCTGACACTTGTCTGATCTCGGACTTCGGAGTGGCAGCAGGTAGGAACA		1386
Db	491	AlaValSerLeuThrLeuAlaValLeuLeuGlyLeuGlyIleThrAlaGlyIleGlyThr		510
Qy	1387	GGAAACAGCTGCCCTGGTCACGGGACACAGCAGCTAGAAAACAGACTTAGTAACATCAT		1446
Db	511	GlySerThrAlaLeuIleLysGlyProIleAspLeuGlnGlnGlyLeuThrSerLeuGln		530
Qy	1447	CGAATTGTACAGAAGATCTCCAAAGCCCTAGAAAATCTGTCTAGTAACCTGGAGGAATCC		1506
Db	531	IleAlaIleAspAlaAspLeuArgAlaLeuGlnAspSerValSerLysLeuGluAspSer		550
Qy	1507	CTAACCTCCTTATCTGAAGTAGTCCCTACAGAATAGAGAGGTTAGATTTATTATTCTA		1566
Db	551	LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu		570
Qy	1567	AAAGAAGAGGATTATGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTCGATCATTC		1626
Db	571	LysGluGlyGlyLeuCysAlaAlaLeuLysGluGlyCysPheTyrIleAspHisSer		590
Qy	1627	GGGGCCATCAGAGACTCCATGACACAGCTTAGAGAAAGTTTGGAGAGCGTCGAGGGAA		1686
Db	591	GlyAlaValArgAspSerMetLysLysLeuLysGluLysLeuAspLysArgGlnLeuGlu		610
Qy	1687	AAGCAACTACTCAAGGTTGGTTTGGAGGATGGTTTCAACAGGCTCTCTTCTGGTGGCTACC		1746
Db	611	ArgGlnLysSerGlnAsnTrpTyrGluGlyTrpPheAsnAsnSerProTrpPheThrThr		630
Qy	1747	CTACTTTCTGCTTTTAACAGGACCCCTTAATAGTCTCTCTCTGTTTACTCACAGTTGGGCCA		1806

```

||||| :: |||||||:||||| |||||||:||||| |||||||:||||| |||||||:|||||
631 LeuLeuSerThrIleAlaGlyProLeuLeuLeuLeuLeuLeuLeuLeuLeuGlyPro 650

1807 TGTATTATTAAACAGTTTAATTCCTTCATTAGAGAACGAATAAGTGCA 1854
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
651 CysIleIleAsnLysLeuValGlnPheIleAsnAspArgIleSerAla 666

RESULT 12
US-09-075-272-5
; Sequence 5, Application US/09075272
; Patent No. 6136598
; GENERAL INFORMATION:
; APPLICANT: MILLER, A. DUSTY
; APPLICANT: WOLGAMOT, GREG
; APPLICANT: BONEHAM, LYNN
; TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL
; TITLE OF INVENTION: PACKAGING CELL LINES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/075,272
; FILING DATE: 08-MAY-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/046,140
; FILING DATE: 09-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Poof, Brian W.
; REGISTRATION NUMBER: 32,928
; REFERENCE/DOCKET NUMBER: 14538A-003710
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-075-272-5

```

Alignment Scores:	2.56e-140	Length:	673
Pred. No.:	1477.50	Matches:	327
Score:		Conservative:	78
Percent Similarity:	59.0%	Mismatches:	161
Best Local Similarity:	47.6%	Indels:	121
Query Match:	42.6%	Gaps:	26
DB:	2		

```
QY 235 -----CCTTAGGAACCTGGTGGCTGATCTATAC-----GTTTCG---CTCAGA 276
Db 61 AlaValHisProProTrrPThrTrrPProAspLeuThrProAspIleCysLeuLeuAla 80
QY 277 TCAGTTATTCTCTAGT-----CTGACCTCACCCCGAGAT--- 309
Db 81 AlaGlySerProAsnTrrPAspLeuProAspHisThrAspLeuAsnAsnProProSerGlu 100
QY 309 ----- 309
Db 101 GlnLysCysValProAsnGlyValGlySerThrThrGlyCysSerGlyGlnPheTyrArg 120
QY 310 ---ATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACACCAATAATAGGAAAA--- 363
Db 121 AlaAsnLeuArgAlaAlaGlnPheTyrValCysProGlyGlnGlyGlnLysGlyLysLeu 140
QY 364 -----CATTCGCGAATCCAGAGATTTCTTTGTAAACAATGGAACCTGTGTAACCTCT 417
Db 141 GlnGlnGluCysArgGlyAlaSerAspTyrPheCysGlyLysTrrPThrCysGluThrThr 160
QY 418 AATGATGATATTGAAATGCCAACCTCTCAGCAGGATAGGTAAGTTTCTTATGTC 477
Db 161 GlyGluAlaTyrTrpLys---ProSerAlaAspTrpAspLeuIleThrValLysArgGly 179
QY 478 AACACCTAT-----ACGACTCTGGACAATTTAATACCTGACTCGACTGGATAGAACTGGA 531
Db 180 SerGlyTyrAspLysProAsnGlnGlyGluArgAsnProTyrLysTyrLeuAspSerGly 199
QY 532 -----AGCCCCAAG-----TGCTCTCCCTCA 552
Db 200 CysAlaLeuLysAsnTyrSerProGlyProCysLysGlyLysTyrCysAsnPro--- 218
QY 553 GACCTAGATTACCTAAAAAATAGTTTCACTGAGAAAGGAAACCAAGAAAATATCTTAA 612
Db 219 -----LeuLeuIleLysPheThrGluLysGlyLysGlnAlaArg---LeuSer 233
QY 613 TGGGTAAATGATGTCTTGGGAAATGGTATATTATGGAGCTCGGTAAACACACAGGC 672
Db 234 TrpLeuLysGlyAsnArgTrrPThrArgValTrrPLeuProIle---ArgAspProGly 252
QY 673 TCACATTTAATCTATTCGCTCAAAATAAACCCAGCTGGAGCTCAATGGCTATAGGACCA 732
Db 253 PheIlePheThrIleArgLeuThrValArgAspLeu---AlaValThrSerIleGlyPro 271
QY 733 AATACGCTCTTGACGGGTCAA----- 753
Db 272 AsnLysValLeuThrGluGlnAlaProProValAlaProAlaProAlaProArgValProAla 291
QY 754 -----AGACCCCAACCCAGGACCA-----GGACCATCTCTTAACATAACTTCT 798
Db 292 ValProAlaProProThrSerArgProTrrPThrValGlyProSerLeuGluThrThrLeu 311
QY 799 GGATCAGACCCC-----ACTGACTTAAACGACGACGACTAAATGGGGGCAAACTT 849
Db 312 AlaSerProProLeuLeuAspThrGluAsn-----ArgLeu 323
QY 850 TTTAGCTCTCAGGAGCTTTTCAAGCTCTTAACCTCCAGCTCCAGAGGCTACCTCT 909
Db 324 ValSerLeuValGlnGlyAlaPheLeuValLeuAsnArgThrAsnProAsnMetThrGln 343
QY 910 TCTTTGTTGGCTATGCTTAGCTTCGGGCCACCTTACTATGAAGAAATGGCTGAAGAGGG 969
Db 344 SerCysTrpLeuCysTyrAlaSerAsnProProTrrPThrGluGlyIleAlaGlnThrArg 363
QY 970 AATTCATGTGCAAAAGAACATAGAGACAATGCATGGGATCCCAAAATAAGCTT 1029
Db 364 ThrTyrAsnIleThrSerAspHis---SerGlnCysLeuTrrPThrGluAsnArgLysLeu 382
QY 1030 ACCCTTACTGAGTTCTCGAAAGGACCTGCAAGAAAGTTCCCTCCATCCACCAA 1089
Db 383 ThrLeuThrAlaValSerGlyAsnGlyLeuCysLeuGlyGlnValProGlnAspLysTrrP 402
```

```
QY 1090 CACCTTTGTAACACACAGTGAAGCCCTTTAATCAAACTCTGAGAGTCAATATCTGTACCT 1149
Db 403 HisLeuCysAsnGlnThrGlnAsnIleArgProAsnLysGlyGlyGlnTrrPLeuValPro 422
QY 1150 GGTATTACACAGGTGGTGGCATGTAATACCTGATTAACCCCTTGTGTTCCACCTTGTT 1209
Db 423 ProIleAspThrValTrrPAlaCysAsnThrGlyLeuThrProCysIleSerMetSerVal 442
QY 1210 TTTAACCAAACTAAAGATTTTTCATATGCTCCAAATTTGTTCCCGAGTGTATTACTAT 1269
Db 443 PheAsnSerSerLysAspPheCysIleLeuValGlnLeuIleProArgLeuLeuTrrHis 462
QY 1270 CCCGAAAAGCAATCCTTGTATGATATGACTACAGAAATCATCAGCAAAAGAGAGAACCC 1329
Db 463 AspAspSerSerPheLeuAspLysPheGluHisArg---ValA-rgTrrPLeuArgGluPro 481
QY 1330 ATATCTCTGACACTGTCTGATGCTCGACTCGAGTGGAGTG---GCAGGAGGTGAGGACAA 1386
Db 482 IleThrLeuThrLeuAlaValLeuLeuGlyLeuGlyValAlaAlaIleGlyValGlyThr 501
QY 1387 GGAACAGCTGCCCTGTGTCACGGGACACAGCAGCTAGAAAACAGGACTTAGTAACCTACAT 1446
Db 502 GlyThrAlaAlaLeuIleGlnThrProArgTrrPheGlu-----GluLeuArg 517
QY 1447 CGAATTGTAACAGAGATCTCAAGCCCTAGAAAATCTGTGACTAACCCTGGAGAAATCC 1506
Db 518 ThrAlaMetAspThrAspLeuArgAlaIleGluHisSerIleThrLysLeuGluGluSer 537
QY 1507 CTAACCTCCTTATCTGAAGTAGTCTCTACAGAAATAGAGAGGTGTAGATTATTATTCTTA 1566
Db 538 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu 557
QY 1567 AAAGAACGAGGATTTATGTAGCTTCAAGGAGAAATGCTGTTTTTTATGTGATCATTTCA 1626
Db 558 LysGluGlyGlyLeuCysAlaAlaLeuLysGluGluCysPheTrrPValAspHisSer 577
QY 1627 GGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAAGTTGGAGAACGCTCGAAGGAA 1686
Db 578 GlyValIleLysAspSerMetAlaLysLeuArgGluArgLeuAspIleArgGlnArgGlu 597
QY 1687 AAGAAACTACTCAAGGTGTTTGGAGATGTTTCAACAGGTCTCTTTGTTGGCTACCC 1746
Db 598 ArgGluSerLysGlnGlyTrrPheGluSerTrrPheAsnLysSerProTrrPLeuThrThr 617
QY 1747 CTACTTTCTCTTAAACAGGACCTTAATAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1806
Db 618 LeuLeuSerThrIleAlaGlyProLeuIleIleLeuLeuLeuLeuLeuLeuThrPheGlyPro 637
QY 1807 TGTATTATTAAAGTTAATTTGCTTTCATTTAGAGAACGAATTAAGTGCAGTCCAGATCATG 1866
Db 638 CysIleLeuAsnLysLeuValAlaPheIleArgGluArgIleAsnAlaValGlnValMet 657
QY 1867 GTACTTAGACAACAGTACCAA 1887
Db 658 ValLeuLysGlnTrrPLeuGln 664
```

## RESULT 13

```
US-09-309-572-14
; Sequence 14, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 665
; TYPE: PRT
```

; ORGANISM: Moloney murine leukemia virus

; FEATURE:

; OTHER INFORMATION: env protein

US-09-309-572-14

## Alignment Scores:

Pred. No.: 2,75e-124 Length: 665  
 Score: 1319.00 Matches: 291  
 Percent Similarity: 55.7% Conservative: 86  
 Best Local Similarity: 43.0% Mismatches: 198  
 Query Match: 38.0% Indels: 102  
 DB: 2 Gaps: 20

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-09-309-572-14 (1-665)

QY 49 CCGAAAAGACGTAAGTCCCTTAAAGTTCGCTCCATCGCGTGGTTCCTTACTCTGTCA 108  
 |||||  
 Db 16 ProArgGlyProLeuIleProLeuIleLeuMetLeuArgGlyValSerThrAlaSer 35  
 |||||  
 QY 109 ATACCTCTCAGACTAAGTGTATGCGCATAGGACACGCTGAACCTCCATAAACCCCTTA 168  
 |||||  
 Db 36 ProGlySer-----SerProHisGlnValTyr 44  
 |||||  
 QY 169 TCTCTACCTGTTAATTACTGACTCCGCGACAGGTATTATATCAACACACTCAAGGG 228  
 |||||  
 Db 45 AsnIleThrTrpGluValThr---AsnGlyAspArgGluThrValTrpAlaThrSerGly 63  
 |||||  
 QY 229 GAGGCTCCTTAGGAACCTGTGGCTGATCTATAC-----GTTGCTC----- 273  
 |||||  
 Db 64 AsnHisProLeuTrpTrpProAspLeuThrProAspLeuCysMetLeuAlaHis 83  
 |||||  
 QY 274 -----AGATCAGTATTCTAGTCTGACCTCACCC 303  
 |||||  
 Db 84 HisGlyProSerTyrTrpGlyLeuGluTrpGlnSerProPheSerProGlyPro 103  
 |||||  
 QY 304 CCA----- 306  
 |||||  
 Db 104 ProCysCysSerGlyGlySerSerProGlyCysSerArgAspCysGluGluProLeuThr 123  
 |||||  
 QY 307 -----GATATCTCCAT 318  
 |||||  
 Db 124 SerLeuThrProArgCysAsnThrAlaTrpAsnArgLeuLeuLeuAspGlnThrHis 143  
 |||||  
 QY 319 GCT-----CAGCGATTTATGTTGCGCCAGACCA-----CCAAATAATCGAAACAT 366  
 |||||  
 Db 144 LysSerAsnGluGlyPheTyrValCysProGlyProHisArgProArgGluSerLysSer 163  
 |||||  
 QY 367 TCGCGAAATCCAGAGATTTCTTTGTAAACAATGGAACCTGTGAACCTCTTAATGATGGA 426  
 |||||  
 Db 164 CysGlyGlyProAspSerPheTyrCysAlaTyrTrpGlyCysGluThrThrGlyArgAla 183  
 |||||  
 QY 427 TATTGGAATGGCCAACTCTCAGCAGATAGGGTAAGTTTCTTATGTCAACACCTTAT 486  
 |||||  
 Db 184 TyrTrpLys---ProSerSerSerTrpAsp-----PheIleThrValAsnAsnAsn 199  
 |||||  
 QY 487 ACCAGCTCTGACAAATTTAATTACCTGACCTGGATAGACTGGAGCCCAAGTGCTCT 546  
 |||||  
 Db 200 LeuThrSerAspGln-----AlaValGlnValCysLysAspAsnLysTrpCysAsn 216  
 |||||  
 QY 547 CTTTCAGACCTAGATTACCTAAATAAGTTTCACTGAGAAAGGAAACAAAGAAATATC 606  
 |||||  
 Db 217 Pro-----LeuValIleArgPheThrAspAlaGlyArgArg-----Val 229  
 |||||  
 QY 607 CTAATAATGGTAAATGGTATGTTCTGGGGAATGGTATATTATGGAGGCTCGGGTAAACAA 666  
 |||||  
 Db 230 ThrSerTrpThrThrGlyHisTyrTrpGlyLeuArgLeuTyr---ValSerGlyGlnAsp 248  
 |||||  
 QY 667 CCAGGCTCCATTCTAATCTATTGCGCTCAAAATAAACAGCTGGAGCCCTCAATGGCTATA 726  
 |||||  
 Db 249 ProGlyLeuThrPheGlyIleArgLeuArgTyrGlnAsnLeuGlyProArgValProIle 268  
 |||||  
 QY 727 GGACCAAAATACGGTCTTCAGCGGTCAAAGACCC-----CCAAACCAAGGACCAAGGA 777  
 |||||

Db 269 GlyProAsnProValLeuAlaAspGlnProLeuSerLysProLysProValLysSer 288  
 |||||  
 QY 778 CCATCCTCTACATAACTTCTGGATCAGACCCCTGAGTCTTAACAGACACACTAAATAG 837  
 |||||  
 Db 289 ProSerValThrLysProProSerGlyThrProLeuSerProThrGlnLeuProProAla 308  
 |||||  
 QY 838 GGGGCA-----AAACTTTTGTAGCTCATCCAGGAGGCTTTTCAAGTCTTAACTCCACG 891  
 |||||  
 Db 309 GlyThrGluAsnArgLeuLeuAsnLeuValAspGlyAlaTyrGlnAlaLeuAsnLeuThr 328  
 |||||  
 QY 892 ACTCAGAGGCTACCTCTTCTGTGTGTAGCTATGCTTCCGGCCCACTTACTATGAA 951  
 |||||  
 Db 329 SerProAspLysThrGlnGluCysTrpLeuValAlaGlyProProTyrTyrGlu 348  
 |||||  
 QY 952 GGAATGGCTAGACAGAGGAAATTCATCTGCAAAAAGACATAGAGACCATGCACATGG 1011  
 |||||  
 Db 349 GlyValAlaValLeuGlyThrTyrSerAsnHisThrSerAlaProAlaAsnCysSerVal 368  
 |||||  
 QY 1012 GGATCCCAATAAGCTTACCTTACTGAGGTTCTGGAAAAGGCACCTGCATAGGAAAG 1071  
 |||||  
 Db 369 AlaSerGlnHisLysLeuThrLeuSerGluValThrGlyGlnGlyLeuCysIleGlyAla 388  
 |||||  
 QY 1072 GTTCCCCCATCCCAACACACTTTGTAAACACACTGAAGCCTTTAATCAAACTCTGAG 1131  
 |||||  
 Db 389 ValProLysThrHisGlnAlaLeuCysAsnThrThr-----GlnThrSerSer 404  
 |||||  
 QY 1132 -----AGTCATATCTGGTACCTGGTTATGACAGGTGGTGGGATGTAATACTGGATTA 1185  
 |||||  
 Db 405 ArgGlySerTyrTyrLeuValAlaProThrGlyThrMetTrpAlaCysSerThrGlyLeu 424  
 |||||  
 QY 1186 ACCCTTGTGTTCACCTGCTGTTTAAACCAAAAGATTGTCATTATGTCCTCAA 1245  
 |||||  
 Db 425 ThrProCysIleSerThrThrIleLeuAsnLeuThrThrAspTyrCysValLeuValGlu 444  
 |||||  
 QY 1246 ATTGTTCCCGAGTGTATTACTATCCGAAAAGCAATCTTGTATGAATATGACTACAGA 1305  
 |||||  
 Db 445 LeuTrpProArgValThrTyrHisSerProSerTyrValTyrGlyLeuPheGlu---Arg 463  
 |||||  
 QY 1306 AATCATCGCAAAAAGAGAGAACCCATATCTCTGACACTTGTGCTGTGCTCGGA----- 1359  
 |||||  
 Db 464 SerAsnArgHisLysArgGluProValSerLeuThrLeuAlaLeuLeuGlyGlyLeu 483  
 |||||  
 QY 1360 -----CTTGAGTGGCAGCAGGTGTAGGAACAGCAACAGCTGCGCTGGTCAACGGACCA 1413  
 |||||  
 Db 484 ThrMetGlyGlyIleAlaAlaGlyIleGlyThrGlyThrAlaLeuMetAlaThrGln 503  
 |||||  
 QY 1414 CAGCAGCTAGAAACAGGACTTAGTAACCTACATCGAATTTGTAACAGAGATCTCCAAAGCC 1473  
 |||||  
 Db 504 Gln-----PheGlnGlnLeuGlnAlaValGlnAspAspLeuArgGlu 518  
 |||||  
 QY 1474 CTAGAAAATCTGTCAGTAACCTGGAGAAATCCCTAACCTCTTATCTGAAGTAGTCCCTA 1533  
 |||||  
 Db 519 ValGluLysSerIleSerAsnLeuGluLysSerLeuThrSerLeuSerGluValValLeu 538  
 |||||  
 QY 1534 CAGAAATGAGAGGTTAGATTATTATTTCTAAAAGAGGAGGATTATGTGTAGCCCTG 1593  
 |||||  
 Db 539 GlnAsnArgArgGlyLeuAspLeuLeuPheLeuLysGluGlyLeuCysAlaAlaLeu 558  
 |||||  
 QY 1594 AAGGAGGAATCTGTTTTTATGTGATCATTTCCAGGGCCATCAGAGACTCCATGAACAAAG 1653  
 |||||  
 Db 559 LysGluGluCysCysPheTyrAlaAspHisThrGlyLeuValArgAspSerMetAlaLys 578  
 |||||  
 QY 1654 CTTAGAGAAAGTTGGAGACGCTCGAAGGAAAGAAAGAACTACTCAAGGGTGGTTGAG 1713  
 |||||  
 Db 579 LeuArgGluArgLeuAsnGlnArgGlnLysLeuPheGluSerThrGlnGlyTrpPheGlu 598  
 |||||  
 QY 1714 GGATGTTTCAACAGTCTCTTTGTTGGTACCTCTCTCTCTTCTGCTTTTAAACAGGACCCCTTA 1773  
 |||||  
 Db 599 GlyLeuPheAsnArgSerProTrpPheThrThrLeuIleSerThrIleMetGlyProLeu 618  
 |||||  
 QY 1774 ATAGTCTCTCTCTCTTACTCAGAGTTGGGCGCATGTATTATTAAAGTGAATTCCTTC 1833  
 |||||  
 Db 619 IleValLeuLeuMetIleLeuLeuPheGlyProCysIleLeuAsnArgLeuValGlnPhe 638







Db 1052 ArgGlySerTyrTyrLeuValAlaProThrGlyThrMetTyrAlaCysSerThrGlyLeu 1071  
Qy 1186 ACCCTTGTGTTCCACCTGGTTTTTAACCAACTAAAGATTTTTGCAATTATGTCCTCAA 1245  
Db 1072 ThrProCysIleSerThrThrIleLeuAsnLeuThrThrAspTyrCysValLeuValGlu 1091  
Qy 1246 ATTGTTCCCGAGTGTATTACTATCCCGAAAAGCAATCCTTGATGATATATGACTACAGA 1305  
Db 1092 LeuTyrProArgValThrTyrHisSerProSerTyrValTyrGlyLeuPheGlu---Arg 1110  
Qy 1306 AATCATCGACAAAGAGAGAACCCATATCTCTGACACTTGCTGTGATGCTCGGA----- 1359  
Db 1111 SerAsnArgHisLysArgGluProValSerLeuThrLeuAlaLeuLeuLeuGlyGlyLeu 1130  
Qy 1360 -----CTTGAGTGGCAGCGGTGTAGGAACAGGAACAGCTGCCCTGGTCACGGGACCA 1413  
Db 1131 ThrMetGlyGlyIleAlaAlaGlyIleGlyThrGlyThrAlaLeuMetAlaThrGln 1150  
Qy 1414 CAGCAGCTAGAAACAGGACTTAGTAACCTACATCGAATTGTAAACAGAAGATCTCCAGGCC 1473  
Db 1151 Gln-----PheGlnGlnLeuGlnAlaAlaValGlnAspSerLeuArgGlu 1165  
Qy 1474 CTAGAAAATCTGTGAGTAACCTGGAGGAATCCCTTAACCTCTTATCTGAAGTAGTCCTA 1533  
Db 1166 ValGluLysSerIleSerAsnLeuGluLysSerLeuThrSerLeuSerGluValValLeu 1185  
Qy 1534 CAGAAATAGAGAGGTTAGATTTATTATTTCTTAAAGAGGAGGATTATGTAGCCTTG 1593  
Db 1186 GlnAsnArgArgGlyLeuAspLeuPheLeuLysGluGlyLeuCysAlaAlaLeu 1205  
Qy 1594 AAGAGGAATGCTGTTTATGTGATTCATTTCAGGGGCCATCAGAGCTCCATGAACAAG 1653  
Db 1206 LysGluGluCysCysPheTyrAlaAspHisThrGlyLeuValArgAspSerMetAlaLys 1225  
Qy 1654 CTTAGAGAAAGTTGGAGACGCTCGAAGGGAAGAAACTACTCAAGGGTGGTTGAG 1713  
Db 1226 LeuArgGluArgLeuAsnGlnArgGlnLysLeuPheGluSerThrGlnGlyTyrPheGlu 1245  
Qy 1714 GGATGTTTCAACAGGTCTCTTTGGTGGCTACCTTCTTCTTAAACAGGACCCCTTA 1773  
Db 1246 GlyLeuPheAsnArgSerProTyrPheThrThrLeuIleSerThrIleMetGlyProLeu 1265  
Qy 1774 ATAGTCCTCTCCTGTTACTCACAGTTGGGCCATGTATTATTAACAAGTTAATGCTTC 1833  
Db 1266 IleValLeuLeuMetIleLeuLeuPheGlyProCysIleLeuAsnArgLeuValGlnPhe 1285  
Qy 1834 ATTAGAGACGAATAAGTCAGTCCAGATCATGTGCTACTAGACAAACAGTAC 1884  
Db 1286 ValLysAspArgIleSerValGlnAlaLeuValLeuThrGlnGlnTyr 1302

Search completed: February 14, 2006, 16:18:41

Job time : 98.4616 secs

**This Page Blank (uspto)**

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1489.5	42.9	667	1	VCLJGL	env polypolprotein pr
2	1351	39.0	676	2	S70395	env polypolprotein -
3	1347	38.8	676	2	T01381	env protein - muri
4	1345	38.8	676	1	VCNVPV	env polypolprotein pr
5	1344	38.8	665	1	VCNPKA	env polypolprotein pr
6	1339	38.6	665	1	VCNWEK	env polypolprotein pr
7	1339	38.6	669	1	VCNWEK	env polypolprotein -
8	1335	38.5	689	2	A46511	env polypolprotein -
9	1335	38.5	689	2	B43491	env polypolprotein -
10	1332.5	38.4	688	2	A43491	env polypolprotein -
11	1319	38.0	665	1	VCNWEH	env polypolprotein -
12	1313	37.9	666	1	VCNWHL	env polypolprotein pr
13	1290	37.2	640	1	VCNVRV	env polypolprotein pr
14	1288	37.1	661	1	VCNVCB	env polypolprotein -

Db 11 ThrSerAsnLeuHisLeuArgHisGlnMetSerProGlySerTrpLysArgLeu--- 29  
Qy 64 ATCCCTTAAGCTTCGCCTCCATCCGCTGGTTCCTTACTCTCTCAATACCTCTCAGACT 123  
Db 30 -----llelleLeuLeuSerCysValPheGlyGly 39  
Qy 124 AATGGTATGCGCATAGGAGACAGCCTGAACCTCCCAATAAACCTTATCTCTCACCTGGTTA 183  
Db 40 GlyGlyThrSerLeu---GlnAsnLysAsnProHisGlnProMetThrLeuThrTrpGln 58  
Qy 184 ATTACTGACTCGGCACAGGTATTATATCAACAACACTCAAGGGAGGCTCCTTTAGGA 243  
Db 59 ValLeu---SerGlnThrGlyAspValValTrpAspThrLysAlaValGlnProProTrp 77  
Qy 244 ACCTGGTGGCTGATCTATACGTTTGGCTCAGATCAGTTATTCTAGTCTGACCTCA--- 300  
Db 78 ThrTrpTrpProThrLeuLysProAspValCysAlaLeuAlaAlaSerLeuGluSerTrp 97  
Qy 301 -----CCCCCAGATATC----- 312  
Db 98 AsplleProGlyThrAspValSerSerSerLysArgValArgProProAspSerAspTyr 117  
Qy 312 ----- 312  
Db 118 ThrAlaAlaTyLysGlnlethrTrpGlyAlaIleGlyCysSerTyrProArgAlaArg 137  
Qy 313 -----CTCCATGCTCAGGATTTTATGTTTGGCCAGGACCCCA-----AATAAT 357  
Db 138 ThrArgMetAlaSerSerThrPheTyrValCysProArgAspGlyArgThrLeuSerGlu 157  
Qy 358 GGAAACACATTGGGAAATCCCAAGAGATTCTTTTGTAAACAATGAACCTGTGTAACTCT 417  
Db 158 AlaArgArgCysGlyGlyLeuGluSerLeuTyrCysLysGluTrpAspCysGluThrThr 177  
Qy 418 AATGATGATATG--- 444  
Db 178 GlyThrGlyTyrTrpLeuSerLysSerSerLysAspLeuIleThrValLysTrpAspGln 197  
Qy 445 TCTCAGCAGGATAGGTAAGTTTCTTATGTGTCAACACTATACAGCTCTGGACAAATT 504  
Db 198 AsnSerGluTrpThrGlnLysPheGln-----CysAsnPro----- 208  
Qy 505 AATTACCTGACTCGATTAGAACTGGAAGCCCAAGTCTCTCTCTTCAGACCTAGATTAC 564  
Db 209 HisGlnThrGlyTrp----- 216  
Qy 565 CTAATAAATAGTTTCACTCAGAAAGGAAACAAAGAAATATCTCTAAATGGTAAATGGT 624  
Db 217 LeuLysIleAspPheThrAspLysGlyLysLeuSerLys-----AspTrpIleThrGly 234  
Qy 625 ATGCTCTGGGNAATGGTATATTATGGAGGCTCGGTTAAACAACAGGCTCCATTCTAACT 684  
Db 235 LysThrTrpGlyLeuArgPheTyr---ValSerGly---HisProGlyValGlnPheThr 252  
Qy 685 ATTCCCTCAAAATAAACAGCTGAGCTCCAAATGGCTATAGGACCAAAATACGCTCTTG 744  
Db 253 IleArgLeuLysIleThrAsnMet---ProAlaValAlaValGlyProAspLeuValLeu 271  
Qy 745 ACGGCTCAA-----AGACCCCAACCCCAA 768  
Db 272 ValGluGlnGlyProProArgThrSerLeuAlaLeuProProProProArgGlu 291  
Qy 769 GGACAGGACCATCC-----TCTAAC-----ATACTTCTGGATCAGAC 807  
Db 292 AlaProProProSerLeuProAspSerAsnSerThrAlaLeuAlaThrSerAlaGlnThr 311  
Qy 808 CCCACTGAGCTTAACAGCAGCAGCTAAATG-----GGGCAAAA 846  
Db 312 ProThrValArgLysThrIleValThrLeuAsnThrProProProThrThrGlyAspArg 331  
Qy 847 CTTTTAGCCTCATCAGGAGCTTTTCAAGCTCTTAACCTCCAGACTCCAGAGGCTACC 906  
Db 332 LeuPheAspLeuValGlnGlyAlaPheLeuThrLeuAsnAlaThrAsnProGlyAlaThr 351

Qy 907 TCTTCTTGCTTGGCTATGCTTAGCTTCGGGCCCCACCTTACTATGAGGAATGGCTAGAAGA 966  
Db 352 GluSerCysTrpLeuCysLeuAlaMetGlyProProTyrTyrGluAlaIleAlaSerSer 371  
Qy 967 GGGAAATTCATATGACAAAGAAACATAGAGACCAATGACATGGGGATCCCAAATTAAG 1026  
Db 372 GlyValAlaTyLysSerThrAsp---LeuAspArgCysArgTrpGlyThrGlnGlyLys 390  
Qy 1027 CTTACCTTACTAGGTTTCTGGAAAGGACCTGTCATAGGAAGGTTTCCCCATCCCA 1086  
Db 391 LeuThrLeuThrGluValSerGlyHisGlyLeuCysIleGlyLysValProPheThrHis 410  
Qy 1087 CAACACCTTCTTAACACACACTGAAGCCTTTAATCAAACTCTGAGAGTCAATATCTGTA 1146  
Db 411 GlnHisLysCysAsnGlnThrLeuSerIleAsnSerSerGlyAspHisGlnTrpLeuLeu 430  
Qy 1147 CCTGTTATGACAGCTGGTGGGCATGTAATACTCGATTAAACCCCTGTGTTCACCTTG 1206  
Db 431 ProSerAsnHisSerTrpTrpAlaCysSerThrGlyLeuThrProCysLeuSerThrSer 450  
Qy 1207 GTTTTAAACCAACTAAAGATTTTTCGATTATGTCCTCAAAATGTTTCCCGAGTGTATTAC 1266  
Db 451 ValPheAsnGlnThrArgAspPheCysIleGlnValGlnLeuIleProArgIleTyr 470  
Qy 1267 TATCCGAAAAAGCAATCCTTGATGATATGATGATGATGATGATGATGATGATGATGAT 1326  
Db 471 TyrProGluGluValLeuLeuGlnAlaTyLysAsnSerHisProArgThrLysArgGlu 490  
Qy 1327 CCCATATCTCTGACACTTGTGTGATGTCGGACTTGGAGTGGCAGCAGGTGTAGGAACA 1386  
Db 491 AlaValSerLeuThrLeuAlaValLeuLeuGlyLeuGlyIleThrAlaGlyIleGlyThr 510  
Qy 1387 GGAACAGCTGCCCTGGTCCCGGACCCAGCAGCAGCTAGAGAAACAGGCTTACTAACCTACAT 1446  
Db 511 GlySerThrAlaLeuIleLysGlyProIleAspLeuGlnGlnGlyLeuThrSerLeuGln 530  
Qy 1447 CGAATTGTAAACAGAAAGATCTCAAGCCCTGAGAAATCTGTGAGTAACTCGAGGAATCC 1506  
Db 531 IleAlaIleAspAlaAspLeuArgAlaLeuGlnAspSerValSerLysLeuGluAspSer 550  
Qy 1507 CTAACCTCTCTTATCTGAAGTAGTCTCTCAGAAATAGAGAGGTTAGATTTATTATTCTA 1566  
Db 551 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu 570  
Qy 1567 AAAGAGAGAGATTATGTTAGCTTGAAGAGGAACTCTCTTTTATGTCGATCACTCA 1626  
Db 571 LysGluGlyGlyLeuCysAlaAlaLeuLysGluGlyCysCysPheTyrIleAspHisSer 590  
Qy 1627 GGGGCCATCAGAGACTCCATCAAGCTTAGAGAAAGGTTGGAGAGCGTCGAAGGAA 1686  
Db 591 GlyAlaValArgAspSerMetLysLysLeuLysGluLysLysLysLysLysLysLys 610  
Qy 1687 AAGAAACTACTCAAGGGTGGTTTGGAGGATGGTTCAACAGGTCTCTTTTGGTGGCTACC 1746  
Db 611 ArgGlnLysSerGlnAsnTrpTyrGluGlyTrpPheAsnAsnSerProTrpPheThrThr 630  
Qy 1747 CTACTTTCTGCTTTAAACAGGACCTTAAATAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1806  
Db 631 LeuLeuSerThrIleAlaGlyProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 650  
Qy 1807 TGTATTATTAAACAGTTAATTCCTTCAATTAGAGAACCAATAAGTGCA 1854  
Db 651 CysIleIleAsnLysLeuValGlnPheIleAsnAspArgIleSerAla 666

## RESULT 2

S70395

env polyprotein - Friend murine leukemia virus (strain FB29)

N:Contains: knob protein gp6; R protein; spike protein p15E

C:Species: Friend murine leukemia virus

A:Variety: strain FB29

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004

C:Accession: S70395

R;Perryman, S.; Nishio, J.; Chesebro, B.

Nucleic Acids Res. 19, 6950, 1991

A;Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.

A;Reference number: S70393; MUID:92107687; PMID:1762923

A;Accession: S70395

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: genomic RNA

A;Residues: 1-676 <PER>

A;Cross-references: UNIPROT:P26804; UNIPARC:UPI000012A03C; EMBL:Z11128; NID:G61547; PIDN

A;Experimental source: strain FB29

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991

C;Genetics:

A;Gene: env

C;Superfamily: type C retrovirus env polyprotein

C;Keywords: coat protein; polyprotein; spike protein; transmembrane protein

Alignment Scores:

Pred. No.:	9.56e-100	Length:	676
Score:	1351.00	Matches:	298
Percent Similarity:	55.7%	Conservative:	84
Best Local Similarity:	43.4%	Mismatches:	194
Query Match:	39.0%	Indels:	110
DB:	2	Gaps:	22

US-10-723-552-3\_copy\_5620\_7533 (1-1914) x S70395 (1-676)

Qy	49	CCGAAAGAGACTGAAATCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCA	108
Db	16	ProArgAspLeuLeuLeuProLeuLeuLeu-----PheLeuSerLeu--- 29	
Qy	109	ATAACCTCTCAGACTAATGATGATCGC---ATAGGAGACACCTGAACTCCCATPAAACCC	165
Db	30	-----LysGlyAlaArgSerAlaAlaProGlySerProHisGlnVal 44	
Qy	166	TTATCTCTACCTGGTTAATTAAGTCTCGGCGACAGGTATTATATCAACACACTCAA	225
Db	45	TyrAsnIleThrTrpGluValThr---AsnGlyAspArgGluThrValTrpAlaIleSer 63	
Qy	226	GGGAGAGCTCTTTAGGAACCTGGTGG-----CCTGATCTATACGTTTGGCTC	273
Db	64	GlyAsnHisProLeuThrTrpTrpProValLeuThrProAspLeuCysMetLeuAla 83	
Qy	274	AGATCAGTTATTCCT-----AGTCTGACCTCACCCCCA--- 306	
Db	84	LeuSerGlyProProHisTrpGlyLeuGluTyrGlnAlaProTyrSerSerProGly 103	
Qy	306	----- 306	
Db	104	ProProCysCysSerGlySerSerGlyAsnValAlaGlyCysAlaArgAspCysAsnGlu 123	
Qy	307	-----GAT 309	
Db	124	ProLeuThrSerLeuThrProArgCysAsnThrAlaTrpAsnArgLeuLysLeuAspGln 143	
Qy	310	ATCCTCCAT-----GCTACGGATTTTATGTTTGGCCAGGACCA-----CCAATAAT	357
Db	144	ValThrHisLysSerSerGluGlyPheTyrValCysProGlySerHisArgProArgGlu 163	
Qy	358	GGAAACATCGCGAAATCCAGAGATTCTTTGTAAACAATGGAACTGTAACTCT	417
Db	164	AlaLysSerCysGlyGlyProAspSerPheTyrCysAlaSerTrpGlyCysGluThr 183	
Qy	418	AATGATGATATTGGAATGGCAACCTCTCAGCAGGATAGGGTAAGTTTCTTATGTC	477
Db	184	GlyArgValTyrTrpLys-----ProSerSerSerTrpAspTyrIleThrValAsp----- 200	
Qy	478	AACACCTTATACCACTCTGGACAATTTAATTAATCACTGACCTGGATTTAGAACTGGAAGCCCC	537
Db	201	AsnAsnLeuThrSer-----AsnGlnAlaValGlnValCysLysAspAsnLys 216	
Qy	538	AAGTGTCTCTCAGACCTAGATTACTTAAATAATAGTTTCACTGAGAAGGAACAA	597
Db	217	TrpCysAsnPro-----LeuAlaIleArgPheThrAsnAlaGlyLysGln 231	

Qy	598	GAAATATCTCTAAATGGTAAATGGTATGCTTTGGGAATGGTATATATTAGGAGCTCG	657
Db	232	-----ValThrSerTrpThrGlyHisTyrTrpGlyLeuArgLeuTyr---ValSer 248	
Qy	658	GGTAAACACACGAGCTCCATTCTAACTATTTCGCCTCAAAATAAACACGAGCTGAGGCTCCA	717
Db	249	GlyGlnAspProGlyLeuThrPheGlyLeuArgLeuSerTyrGlnAsnLeuGlyProArg 268	
Qy	718	ATGCTATAGGACCAATACGCTCTTCACGGGTCAA----- 753	
Db	269	IleProIleGlyProAsnProValLeuAlaAspGlnLeuSerPheProLeuProAsnPro 288	
Qy	754	-----AGACCCCAACCCAGGACACGAGCATCTCTTAAC-----ATAACTCT	798
Db	289	LeuProLysProAlaLysSerProProAlaSerSerSerThrProThrLeuLeuSerPro 308	
Qy	799	GGATCAGACCCCTAGTCTTAACAGCAGCACTAAATGGGGGCAAACTTTTATGAGCTC	858
Db	309	SerProThrProThrGlnPro---ProProAlaGlyThrGlyAspArgLeuAsnLeu 327	
Qy	859	ATCCAGGAGCTTTTCAAGCTCTTAACCTCCACGACTCCAGAGGCTACCTCTTCTGTGG	918
Db	328	ValGlnGlyAlaTyrGlnAlaLeuAsnLeuThrAsnProAspLysThrGlnGluCysTrp 347	
Qy	919	CTATGCTTAGCTTCGGGCCCACTTACTATGAAGGAATGGCTAGAGAGGGAATTTCAAT	978
Db	348	LeuCysLeuValSerGlyProProTyrTrpGlyValAlaValLeuGlyThrTyrSer 367	
Qy	979	GTGCAAAAGAACATAGAGACCAATGCATGGGGATCCCAAAATAAGCTTACCTTACT	1038
Db	368	AsnHisThrSerAlaProAlaAsnCysSerValAlaSerGlnHisLysLeuThrLeuSer 387	
Qy	1039	GAGTTTCTCGAAGAGCCTGCATAGGAAGTTCCCCATCCACCCACACACTTGT	1098
Db	388	GluValThrGlyArgGlyLeuCysIleGlyThrValProLysThrHisGlnAlaLeuCys 407	
Qy	1099	AACCACACTGAAGCTTTAATCAAACTCTGAGAGTCAATATCTGGTACCTGGTTATGAC	1158
Db	408	AsnThrThr-----LeuLysAlaGlyLysGlySerTyrTyrLeuValAlaProThrGly 425	
Qy	1159	AGTGGTGGCGATGTAATACTGGATTAAACCTTGTGTTTCCACTTGGTTTAAACAA	1218
Db	426	ThrMetTrpAlaCysAsnThrGlyLeuThrProCysLeuSerAlaThrValLeuAsnArg 445	
Qy	1219	ACTAAAGATTTTGCATTATGTTCCCAATTTGTTCCCGAGCTGATTTACTATCCGAAAA	1278
Db	446	ThrThrAspTyrCysValLeuValGluLeuTrpProArgValThrTyrHisProProSer 465	
Qy	1279	GCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAGAGAGAACCCATATCTG	1338
Db	466	TyrValTyrSerGlnPheGlu---LysSerHisArgHisLysArgGluProValSerLeu 484	
Qy	1339	ACACTTGTCTGTATGCTCGA-----CTTGGAGTGGCAGCAGGTGTAGGAACA	1386
Db	485	ThrLeuAlaLeuLeuLeuGlyGlyLeuThrMetGlyGlyIleAlaAlaGlyValGlyThr 504	
Qy	1387	GGAAACAGCTCCCTGTCACGGGACACACAGCTAGAAACAGGACTTAGTAACCTACAT	1446
Db	505	GlyThrThrAlaLeuValAlaThrGlnGln-----PheGlnGlnLeuHis 519	
Qy	1447	CGAATTGTAACAGAACTCTCCAAAGCCCTAGAAAAATCTGTAGTAACCTGGAGGAATCC	1506
Db	520	AlaAlaValGlnAspLeuLysGluValGluLysSerIleThrAsnLeuGluLysSer 539	
Qy	1507	CTAACCTCTTATCTGAAGTAGTCTCACAAGATAGAGAGGTTTAGATTTATTATTTCTA	1566
Db	540	LeuThrSerLeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeuPheLeu 559	
Qy	1567	AAAGAGGAGGATTATGTGTAGCTTGAAGGAGGAATCTGTTTATGTGGATCATCA	1626
Db	560	LysGluGlyGlyLeuCysAlaAlaLeuLysGluCysCysPheTyrAlaAspHisThr 579	

```
QY 1627 GGGGCCATCAGAGCTCCATCAACAGCTTAGAGAAAGGTTGGAGAGCGTCGAAGGAA 1686
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 580 GlyLeuValArgAspSerMetAlaLysLeuArgGluArgLeuSerGlnArgGlnLysLeu 599
QY 1687 AAGGAACTACTCAAGCGTGGTTGAGGATGGTTCAACAGGTCCTCTTGGTTGGCTACC 1746
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 600 PheGluSerSerGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpPheThrThr 619
QY 1747 CTACTTTCGCTTTAACAGGACCCCTTAATAGTCCTCTCTCTCTCTCTCTCTCTCTCT 1806
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 620 LeuIleSerThrIleMetGlyProLeuIleIleLeuLeuLeuIleLeuPheGlyPro 639
QY 1807 TGTATTATTACAAGTTAATTGGCTTCATTATAGAGAACGAATAAGTGCAGTCCAGATCATG 1866
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 640 CysIleLeuAsnArgLeuValGlnPheValLysAspArgIleSerValValGlnAlaLeu 659
QY 1867 GTACTTAGACAACAGTAC 1884
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 660 ValLeuThrGlnGlnTyr 665

RESULT 3
T01381
env protein - murine leukemia virus
C:Species: murine leukemia virus
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 31-Dec-2004
C:Accession: T01381
R:Turnas, K.M.; Poszgay, J.M.; Avidan, N.; Ksiazek, S.J.; Overmoyer, B.; Blank, K.J.; Pry
Virology 192, 587-595, 1993
A:Title: Loss of antigenic epitopes as the result of env gene recombination in retroviri
A:Reference number: Z14313; MUID:93134803; PMID:7678475
A:Accession: T01381
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-676 <TUM>
A:Cross-references: UNIPROT:O12374; UNIPARC:UPI000017864A; EMBL:S53043; NID:G263651
C:Genetics:
A:Gene: env

Alignment Scores:
Pred. No.: 1.99e-99 Length: 676
Score: 1347.00 Matches: 298
Percent Similarity: 55.4% Conservative: 82
Best Local Similarity: 43.4% Mismatches: 196
Query Match: 38.8% Indels: 110
DB: 2 Gaps: 22

US-10-723-552-3_COPY_5620_7533 (1-1914) x T01381 (1-676)
QY 49 CGGAAAGACTGAAATCCCTTAAAGCTTCGCTCCATCGCGTGGTCTTACTCTGTCA 108
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 16 ProArgAspLeuIleIleProLeuIleLeu 29
QY 109 ATAACCTCTCAGACTAATGGTATGCGC---ATAGGAGACAGCTGAACTCCCATAAACCC 165
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 30 -----LysGlyAlaArgSerAlaAlaProGlySerSerProHisGlnVal 44
QY 166 TTATCTCTCAGCTGTTAATTAAGTCTCCGCGCAGGATTATATATCAACACACTCAA 225
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 45 TyrHisIleThrTrpGluValThr---AsnGlyAspArgGluThrValTrpAlaIleSer 63
QY 226 GGGGAGGCTCCTTTAGGAACCTGGTGG---CCTGATCTATAGCTTTGGCCTC 273
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 GlyAsnHisProLeuTrpTrpProValLeuThrProAspLeuCysMetLeuAla 83
QY 274 AGATCAGTATTTCCT- 300
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 LeuSerGlyProProHisTrpGlyLeuGluTyrGlnAlaProTyrSerSerAlaProGly 103
QY 301 CCCCCA----- 306
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 104 ProProCysCysSerGlySerGlyAsnIleAlaGlyCysAlaArgAspCysAsnGlu 123
QY 307 -----GAT 309
```

```
Db 124 ProLeuThrSerLeuThrProArgCysAsnThrAlaGlyAsnArgLeuLysLeuAspGln 143
QY 310 ATCTCTCAT-----GCTCAGCGATTTTATGTTTGGCCCGCAGGACCA-----CCAAATAAT 357
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 144 ValThrHisLysSerSerGluGlyPheTyrValCysProGlySerHisArgProArgGlu 163
QY 358 GGAACAACATTCGCGAAATCCAGAGATTTCTTTTGTAAACAATCGAACTGTGTAACTCTCT 417
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 164 AlalysSerCysGlyGlyProAspSerPheTyrCysAlaSerTrpGlyCysGluThrThr 183
QY 418 AATGATGATATTGGAAATGGCCAACTCTCAGCAGGATAGGTAAGATTTTCTTATGTC 477
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 184 GlyArgValTyrTrpLys---ProSerSerProTrpAspTyrIleThrValAsp----- 200
QY 478 AACACCTATACCACTCTGGCAATTTAATACCTGACCTGGATAGACTAGGAGAGCCCC 537
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 201 AsnAsnLeuThrSer-----AsnGlnAlaValGlnValCysLysAspAsnLys 216
QY 538 AAGTGCTCTCTTCAGACCTAGATTACTCTAAAATAAGTTTCTACTGAGAAAGGAAACAA 597
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 217 TrpCysAsnPro-----LeuAlaIleGlnPheThrAsnAlaGlyLysGln 231
QY 598 GAAAATATCTAAATGGTAAATGGTATGCTTTGGGGAATGGTATATATTATGGAGCTCG 657
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 232 -----ValThrSerTrpThrThrGlyHisTyrTrpGlyLeuArgLeuTyr---ValSer 248
QY 658 GGTAAACACACAGGCTCCATCTTAACCTATTCGCTCAAAATAAACACCACTGGAGCTTCCA 717
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 GlyGlnAspProGlyLeuThrPheGlyIleArgLeuSerTyrGlnAsnLeuGlyProArg 268
QY 718 ATGGCTATAGACCAATAACGCTCTTGACGGGTCAA----- 753
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 269 IleProIleGlyProAsnProValLeuAlaAspGlnLeuSerPheProLeuProAsnPro 288
QY 754 -----AGACCCCAACCCAGGACGAGGACCATCTCTCTAAC-----ATACTTCT 798
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 289 LeuProLysProAlaLysSerProAlaLysSerSerSerSerThrProThrLeuIleSerPro 308
QY 799 GGATCAGACCCCACTAGCTTAACAGCAGCACTAAATGGGGGCAAACTTTTAGCCTC 858
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 309 SerProThrProThrGlnPro---ProProValGlyThrGlyAspArgLeuLeuAsnLeu 327
QY 859 ATCCAGGAGGCTTTTCAAGCTCTTAACCTCCAGCACTCCAGAGGCTACCTCTTCTGTGG 918
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 328 ValGlnGlyAlaTyrGlnAlaLeuAsnLeuThrAsnProAspLysThrGlnGluCysTrp 347
QY 919 CTATGCTTAGCTCGGGCCCACTTACTATCAAGAAGATGGCTAGAGAAGGAAATTCAT 978
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 348 LeuCysLeuValSerGlyProProTyrTyrGluGlyValAlaValLeuGlyThrTyrSer 367
QY 979 GTGCAAAAGAACATAGACCAATGCACATCGGGATCCCAATTAAGCTTACCTTACT 1038
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 368 AsnHisThrSerAlaProAlaAsnCysSerValAlaSerGlnHisLysLeuThrLeuSer 387
QY 1039 GAGGTTCTGAAAGGACCTGCATAGGAAGGTTCCCCATCCCAACCAACACCTTGT 1098
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 388 GluValThrGlyArgGlyLeuCysIleGlyThrValProLysThrHisGlnAlaLeuCys 407
QY 1099 AACCACATCAAGCCTTTAATCAAACTCTCGAGAGTCAATATCTGGTACCTGGTTATGAC 1158
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 408 AsnThrThr-----LeuLysAlaGlyLysGlySerTyrTyrLeuValAlaProThrGly 425
QY 1159 AGGTGGTGGGATGATATACCTGGATTAACCCCTTGTGTGTTCCACCTCGTTTAAACCA 1218
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 426 ThrMetTrpAlaCysAsnThrGlyLeuThrProCysLeuSerAlaThrValLeuAsnArg 445
QY 1219 ACTAAAGATTTTGCATTATGTCCTCAAAATGTTCCCGAGTGTATTACTATCCCGAAAAA 1278
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 446 ThrThrAspTyrCysValLeuValGluLeuTrpProArgValThrTyrHisProSer 465
QY 1279 GCAATCTCTTGATGATATGACTACAGAAATCATCGACAAAGAGAGAACCCATATCTCTG 1338
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 466 TyrValTyrSerGlnPheGlu---LysSerHisArgHisLysArgGluProValSerLeu 484  
 QY 1339 ACACCTTGCTGTGATCTCGGA-----CTTGAGTGGCAGCAGGTTGAGGAACA 1386  
 Db 485 ThrLeuAlaLeuLeuGlyGlyLeuThrMetGlyGlyIleAlaAlaGlyValGlyThr 504  
 QY 1387 GGAACAGCTGCCCTGCTCAGCGGACCAAGCAGCTAGAAACAGAGCTTAGTAACCTACAT 1446  
 Db 505 GlyThrThrAlaLeuValAlaThrGlnGln-----PheGlnGlnLeuHis 519  
 QY 1447 CGAATTGTAAACAGAGATCTCCAGCCCTAGAAAATCTGTCAGTAACTCGAGGAAATCC 1506  
 Db 520 AlaAlaValGlnAspAspLeuLysGluValGlnLysSerIleThrAsnLeuGluLysSer 539  
 QY 1507 CTAACCTCTCTTATCTGAAGTAGTCTCTACAGAAATAGAGAGGTTAGATTTATTATTCTA 1566  
 Db 540 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuPheLeu 559  
 QY 1567 AAAGACGAGGATTTATGTGTAGCTTGAAGGAGGAATGCTGTTTTATGTGGATCATCA 1626  
 Db 560 LysGluGlyGlyLeuCysAlaAlaLeuLysGluCysCysPheTyrAlaAspHisThr 579  
 QY 1627 GGGGCCATCAGAGCTCCATCAACAAAGCTTAGAGAAAGTTGGAGAGCGTCGAAGGGA 1686  
 Db 580 GlyLeuValArgSerMetAlaLysLeuArgGluArgLeuSerGlnArgGlnLysLeu 599  
 QY 1687 AAGAAACTACTCAAGGGTGGTTTGGAGGATGGTTCAACAGCTCTCTTTGGTGGCTACC 1746  
 Db 600 PheGluSerSerGlnGlyTyrPheGluGlyTyrPheAsnArgSerProTyrPheThrThr 619  
 QY 1747 CTACTTCTGCTTTAAACAGGACCTTAATAGTCTCTCTCTGTTTACTCAGATGGGCCCA 1806  
 Db 620 LeuLeuSerThrIleMetGlyProLeuLeuIleLeuLeuLeuLeuLeuPheGlyPro 639  
 QY 1807 TGATTATTAAACAGTAATTCCTTCATTAGAGAAAGTAAGTCAAGTCCAGATCATG 1866  
 Db 640 CysIleLeuAsnArgLeuValGlnPheValLysAspArgIleSerValValGlnAlaLeu 659  
 QY 1867 GTACTTAGACAACAGTAC 1884  
 Db 660 ValLeuThrGlnGlnTyr 665  
 RESULT 4  
 VCMVPV  
 env polyprotein precursor - Friend murine leukemia virus (strain PVC-211)  
 N;Alternate names: coat polyprotein  
 N;Contains: knob protein gp76; R protein; spike protein p15E  
 C;Species: Friend murine leukemia virus  
 C;Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 09-Jul-2004  
 C;Accession: A38210; S35476  
 R;Masuda, M.; Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.  
 J. Virol. 66, 2798-2806, 1992  
 A;Title: Molecular characterization of a neuropathogenic and nonerythrocytogenic vari  
 A;Reference number: A38210; MUID:92219364; PMID:1560524  
 A;Accession: A38210  
 A;Molecule type: DNA  
 A;Residues: 1-676 <SPE>  
 A;Cross-references: UNIPROT:P26803; UNIPARC:UPI000012A03D; GB:M931134; NID:g331898; PIDN:  
 R;Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.; Masuda, M.  
 Nucleic Acids Res. 20, 3249, 1992  
 A;Title: Complete nucleotide sequence of a neuropathogenic variant of Friend murine leuk  
 A;Reference number: S35474; MUID:92319660; PMID:1620621  
 A;Accession: S35476  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-676 <REM>  
 A;Cross-references: UNIPARC:UPI000012A03D; EMBL:M931134; NID:g331898; PIDN:AAA46478.1; PT  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992  
 C;Genetics:  
 A;Gene: env  
 C;Superfamily: type C retrovirus env polyprotein  
 C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei  
 F;1-35/Domain: signal sequence #status predicted <SIG>

F;36-676/Product: env polyprotein #status predicted <ENV>  
 F;36-619/Domain: extracellular #status predicted <EXT>  
 F;36-479/Product: knob protein gp76 #status predicted <KGP>  
 F;476-479/Region: cleavage processing #status predicted  
 F;480-659/Product: spike protein p15E #status predicted <SP>  
 F;486-502/Region: hydrophobic #status predicted  
 F;620-636/Domain: transmembrane #status predicted <TM>  
 F;637-676/Domain: intracellular #status predicted <INT>  
 F;660-676/Product: R protein #status predicted <RPT>  
 F;46, 202, 336, 368, 408, 444/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:  
 Pred.: No.: 2.88e-99 Length: 676  
 Score: 1345.00 Matches: 297  
 Percent Similarity: 55.6% Conservative: 86  
 Best Local Similarity: 43.1% Mismatches: 190  
 Query Match: 38.8% Indels: 116  
 DB: 1 Gaps: 22

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x VCMVPV (1-676)

QY 49 CCGAAAAGACTGAAATCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTC 108  
 Db 16 ProArgAspLeuLeuIleProLeuLeu-----PheLeuSerLeu--- 29  
 QY 109 ATAACCTCTCAGACTAATGGTATGCGC---ATAGGAGACAGCCTGAACTCCCATAAACCC 165  
 Db 30 -----LysGlyAlaArgSerAlaAlaProGlySerSerProHisGlnVal 44  
 QY 166 TTATCTCTACCTGGTTAATACTGACTCCGGCAGAGTATTATATCAACAACACTCAA 225  
 Db 45 TyrAsnIleThrTrpGluValThr---AsnGlyAspArgGluThrValTrpAlaIleSer 63  
 QY 226 GGGGAGGCTCTTTAGGAACCTGGTGGCTGATCTATAC-----GTTTGCCTC- 273  
 Db 64 GlyAsnHisProLeuTrpThrTrpProAspLeuThrProAspLeuCysMetLeuAla 83  
 QY 274 -----AGATCAGTTATTCTCTGCTGACCTCA 300  
 Db 84 LeuSerGlyProProHisTrpGlyLeuGluTyrArgAlaProTyrSerSerProGly 103  
 QY 301 CCCCCA----- 306  
 Db 104 ProProCysSerGlySerSerGlyAsnArgAlaGlyCysAlaArgAspCysAspGlu 123  
 QY 307 -----GAT 309  
 Db 124 ProLeuThrSerLeuThrProArgCysAsnThrAlaTrpAsnArgLeuLysLeuAspGln 143  
 QY 310 ATCTCTCCAT-----GCTCAGCGATTATTGTTTGGCCAGACCA-----CCAAATAAT 357  
 Db 144 ValThrHisLysSerSerGlyGlyPheTyrValCysProGlySerHisArgProArgLys 163  
 QY 358 GGAACAATTCGCGAATCCAGAGATTTCTTTGTAACAACATGGAAGTGTGTAACCTCT 417  
 Db 164 AlaLysSerCysGlyGlyProAspSerPheTyrCysAlaSerTrpGlyCysGluThrThr 183  
 QY 418 AATGATGATATTGGAAATGGCCAACTCTCAGCAGATAGGTAAGTTTCTTATGTC 477  
 Db 184 GlyArgAlaTyrTrpLys---ProSerSerSerTyrAspTyrIleThrValAsp- 200  
 QY 478 AACACCTATACACAGCTCTGGACAATTAATTACCTGACCTGGATTAGAACTGGAAGCCCC 537  
 Db 201 AsnAsnLeuThrThr-----AsnGlnAlaGlnValCysLysAspAsnLys 216  
 QY 538 AAGTGTCTCTTTCAGACCTAGATTACCTAAAAAATAGTTTCACTGAGAAGGAAACAA 597  
 Db 217 TrpCysAsnPro-----LeuAlaIleGlnPheThrAsnAlaGlyLysGln 231  
 QY 598 GAAAATATCTAAATGGGTAAATGGTATGCTCTTGGGAATGTTATATTATGAGGCTCG 657  
 Db 232 -----ValThrSerTrpThrIleGlyHisTyrTrpGlyLeuArgLeuTyr---ValSer 248



```
QY 658 GGTAAACACACGAGCTCCATTCTAACTATTGCGCTCAAAATAAACCAGCTGGAGCCTCCA 717
Db 659 GGTAAACACACGAGCTCCATTCTAACTATTGCGCTCAAAATAAACCAGCTGGAGCCTCCA 717
QY 249 GlyGlnAspProGlyLeuThrPheGlyIleArgLeuLysTyrGlnAsnLeuGlyProArg 268
Db 249 GlyGlnAspProGlyLeuThrPheGlyIleArgLeuLysTyrGlnAsnLeuGlyProArg 268
QY 718 ATGGCTATAGGACCAAAATACGGTCTTGAGCGGTCAAGACCC---CCAAACCAAGGACCA 774
Db 718 ATGGCTATAGGACCAAAATACGGTCTTGAGCGGTCAAGACCC---CCAAACCAAGGACCA 774
QY 269 ValProIleGlyProAsnProValLeuAlaAspGlnLeuSerPheProLeuProAsnPro 288
Db 269 ValProIleGlyProAsnProValLeuAlaAspGlnLeuSerPheProLeuProAsnPro 288
QY 775 GGACCATCTCTAACATAACTTCTGGATCAGACCCCACTGAGTCTAACAGCAGCACTAAA 834
Db 775 GGACCATCTCTAACATAACTTCTGGATCAGACCCCACTGAGTCTAACAGCAGCACTAAA 834
QY 289 LeuProLys-----ProAlaLysSerProSerAlaSerAsnSerThrProThr 304
Db 289 LeuProLys-----ProAlaLysSerProSerAlaSerAsnSerThrProThr 304
QY 835 ATG-----GGGCGCAAAACTT 849
Db 835 ATG-----GGGCGCAAAACTT 849
QY 305 LeuLeuSerProSerProAlaProThrGlnProProProAlaGlyThrGlyAspArgLeu 324
Db 305 LeuLeuSerProSerProAlaProThrGlnProProProAlaGlyThrGlyAspArgLeu 324
QY 850 TTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCAGCACTCCAGGCTACTCTCT 909
Db 850 TTTAGCCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCAGCACTCCAGGCTACTCTCT 909
QY 325 LeuAsnLeuValGlnGlyAlaTyrGlnAlaLeuAsnLeuThrAsnProAspLysThrGln 344
Db 325 LeuAsnLeuValGlnGlyAlaTyrGlnAlaLeuAsnLeuThrAsnProAspLysThrGln 344
QY 910 TCTGTGCTGCTATGCTTACGCTTCGGGCCCACTTACTATGAAGNATGGCTAGAGAGGG 969
Db 910 TCTGTGCTGCTATGCTTACGCTTCGGGCCCACTTACTATGAAGNATGGCTAGAGAGGG 969
QY 345 GluCysTrpLeuCysLeuValSerAlaProProTyrTyrGluGlyValAlaValLeuGly 364
Db 345 GluCysTrpLeuCysLeuValSerAlaProProTyrTyrGluGlyValAlaValLeuGly 364
QY 970 AAATTCAATGTGACAAAAGACATAGACCAATGCACATGGGATGCCAAATAAGCTT 1029
Db 970 AAATTCAATGTGACAAAAGACATAGACCAATGCACATGGGATGCCAAATAAGCTT 1029
QY 365 ThrTyrSerAsnHisThrSerAlaProAlaAsnCysSerAlaGlySerGlnHisLysLeu 384
Db 365 ThrTyrSerAsnHisThrSerAlaProAlaAsnCysSerAlaGlySerGlnHisLysLeu 384
QY 1030 ACCCTTACTAGGTTTCTGGAAAAGCCACCTGCATAGGAAAGGTTTCCCCCACTCCACCAA 1089
Db 1030 ACCCTTACTAGGTTTCTGGAAAAGCCACCTGCATAGGAAAGGTTTCCCCCACTCCACCAA 1089
QY 385 ThrLeuSerGluValThrGlyGlnGlyLeuCysIleGlyThrValProLysThrHisGln 404
Db 385 ThrLeuSerGluValThrGlyGlnGlyLeuCysIleGlyThrValProLysThrHisGln 404
QY 1090 CACCTTTGTACACCACTGAAGCCCTTTAATCAAACTCTGAGAGTCAATATCTGGTACCT 1149
Db 1090 CACCTTTGTACACCACTGAAGCCCTTTAATCAAACTCTGAGAGTCAATATCTGGTACCT 1149
QY 405 AlaLeuCysAsnThrThr-----LeuLysThrGlyLysGlySerTyrLeuValAla 422
Db 405 AlaLeuCysAsnThrThr-----LeuLysThrGlyLysGlySerTyrLeuValAla 422
QY 1150 GGTATGACAGCGTGGCGCATGTAATACTGGATTAAACCCCTTGTTGGTTCCACCTTGTT 1209
Db 1150 GGTATGACAGCGTGGCGCATGTAATACTGGATTAAACCCCTTGTTGGTTCCACCTTGTT 1209
QY 423 ProAlaGlyThrMetTrpAlaCysAsnThrGlyLeuThrProCysLeuSerAlaThrVal 442
Db 423 ProAlaGlyThrMetTrpAlaCysAsnThrGlyLeuThrProCysLeuSerAlaThrVal 442
QY 1210 TTTAACCAAACTAAAGATTTTGGCATTATGTTGCCAAATGTTTCCCGAGTGTATPACTAT 1269
Db 1210 TTTAACCAAACTAAAGATTTTGGCATTATGTTGCCAAATGTTTCCCGAGTGTATPACTAT 1269
QY 443 LeuAsnArgThrThrAspTyrCysValLeuValGluLeuTyrProArgValThrTyrHis 462
Db 443 LeuAsnArgThrThrAspTyrCysValLeuValGluLeuTyrProArgValThrTyrHis 462
QY 1270 CCGGAAAAGCAATCTTGATGAATATGACTACAGAAATCATCGACAAAGAGAGAACCC 1329
Db 1270 CCGGAAAAGCAATCTTGATGAATATGACTACAGAAATCATCGACAAAGAGAGAACCC 1329
QY 463 ProProSerTyrValTyrSerGlnPheGlu---LysSerTyrArgHisLysArgGlyPro 481
Db 463 ProProSerTyrValTyrSerGlnPheGlu---LysSerTyrArgHisLysArgGlyPro 481
QY 1330 ATATCTCTGACACTGCTGTGATGCTCGGA-----CTTGGAGTGGCAGCAGGT 1377
Db 1330 ATATCTCTGACACTGCTGTGATGCTCGGA-----CTTGGAGTGGCAGCAGGT 1377
QY 482 ValSerLeuThrLeuAlaLeuLeuLeuLeuGlyLeuThrMetGlyGlyIleAlaAlaGly 501
Db 482 ValSerLeuThrLeuAlaLeuLeuLeuLeuGlyLeuThrMetGlyGlyIleAlaAlaGly 501
QY 1378 GTAGGAACAGGAACAGCTGCCCTTGCTACGGGACACAGCAGCTAGAACAGGACTTAGT 1437
Db 1378 GTAGGAACAGGAACAGCTGCCCTTGCTACGGGACACAGCAGCTAGAACAGGACTTAGT 1437
QY 502 ValGlyThrGlyThrThrAlaLeuValAlaThrGlnGln-----PheGln 516
Db 502 ValGlyThrGlyThrThrAlaLeuValAlaThrGlnGln-----PheGln 516
QY 1438 AACCTACATCGAATGTTGAACAGAGATCTCCAGCCCTAGAAAAATCTGTCAGTAACCTG 1497
Db 1438 AACCTACATCGAATGTTGAACAGAGATCTCCAGCCCTAGAAAAATCTGTCAGTAACCTG 1497
QY 517 GlnLeuHisAlaAlaValGlnAspLeuLysGluValGluLysSerIleThrAsnLeu 536
Db 517 GlnLeuHisAlaAlaValGlnAspLeuLysGluValGluLysSerIleThrAsnLeu 536
QY 1498 GAGGAATCCCTAACTCTTATCTCAAGTAGTCTACAGATAGAGAAGGAGGTTAGATTTA 1557
Db 1498 GAGGAATCCCTAACTCTTATCTCAAGTAGTCTACAGATAGAGAAGGAGGTTAGATTTA 1557
QY 537 GluLysSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 556
Db 537 GluLysSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 556
QY 1558 TTATTTCTAAAGAGGAGGATTTATGTTAGCTTTGAAGGAGGAATGCTGTTTTTATGTG 1617
Db 1558 TTATTTCTAAAGAGGAGGATTTATGTTAGCTTTGAAGGAGGAATGCTGTTTTTATGTG 1617
QY 557 LeuPheLeuLysGluGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrAla 576
Db 557 LeuPheLeuLysGluGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrAla 576
QY 1618 GATCATTCAGGGCCATCAGACATCCATGAACAGCTTAGAAGAAGGTTGAGNAGCGT 1677
Db 1618 GATCATTCAGGGCCATCAGACATCCATGAACAGCTTAGAAGAAGGTTGAGNAGCGT 1677
QY 577 AspHisThrGlyLeuValArgAspSerMetAlaLysLeuArgGluLeuThrGlnArg 596
Db 577 AspHisThrGlyLeuValArgAspSerMetAlaLysLeuArgGluLeuThrGlnArg 596
QY 1678 CGAAGGAAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGTTTCAACAGGTCCTTTGG 1737
Db 1678 CGAAGGAAAAGGAAACTACTCAAGGGTGGTTTGAGGGATGTTTCAACAGGTCCTTTGG 1737
```

```
Db 597 GlnLysLeuPheGluSerSerGlnGlyTrpPheGluGlyLeuPheAsnArgSerProTrp 616
QY 1738 TTGGCTACCTTCTGCTTTTAAACAGGACCCCTTAATAGCTCCTCTCTTACTACTACA 1797
Db 617 PheThrThrLeuIleSerThrIleMetGlyProLeuIleIleLeuLeuLeuIleLeu 636
QY 1798 GTTGGCCCATGTATTATTAAACAAGTAAATTCCTTCTATTAGAGAACGAATAAGTGCAGTC 1857
Db 637 PheGlyProCysIleLeuAsnArgLeuValGlnPheValLysAspArgIleSerValVal 656
QY 1858 CAGATCATGGTACTTACACACAGTAC 1884
Db 657 GlnAlaLeuValLeuThrGlnGlnTyr 665

RESULT 5
VCMVKA
env polyprotein precursor - radiation murine leukemia virus (strain Kaplan)
N;Alternate names: coat polyprotein
N;Contains: coat protein gp76; coat protein p15E; R protein
C;Species: radiation murine leukemia virus
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C;Accession: B42743
R;Poliquin, L.; Bergeron, D.; Fortier, J.L.; Paquette, Y.; Bergeron, R.; Rassart, E.
J. Virol. 66, 5141-5146, 1992
A;Title: Determinants of thymotropism in Kaplan radiation leukemia virus and nucleotide
A;Reference number: A42743; MUID:92333703; PMID:1629969
A;Accession: B42743
A;Molecule type: DNA
A;Residues: 1-665 <POL>
A;Cross-references: UNIPROT:P31794; UNIPARC:UPI000012A042; GB:I93052; NID:G332065; PIDN:
C;Genetics:
A;Gene: env
C;Superfamily: type C retrovirus env polyprotein
C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-665/Product: env polyprotein #status predicted <ENV>
F;32-604/Domain: extracellular #status predicted <EXT>
F;32-467/Product: coat protein gp76 #status predicted <GP1>
F;464-467/Region: cleavage processing #status predicted
F;468-644/Product: coat protein p15E #status predicted
F;474-490/Region: hydrophobic #status predicted
F;605-621/Domain: transmembrane #status predicted <TM1>
F;622-665/Domain: intracellular #status predicted <INT>
F;645-665/Product: R protein #status predicted <RPP>
F;43,199,211,324,356,396,400,432/Binding site: carbohydrate (Asn) (covalent) #status pre

Alignment Scores:
Pred. No.: 3,46e-99 Length: 665
Score: 1344.00 Matches: 92
Percent Similarity: 58.2% Conservative: 297
Best Local Similarity: 44.5% Mismatches: 179
Query Match: 38.8% Indels: 100
DB: 1 Gaps: 21

US-10-723-552-3_copy_5620_7533 (1-1914) x VCMVKA (1-665)
QY 97 CTTACTCTGTCATTAACCTCTCAGACTAATGGTATGGCATAGGACAGCAGCTGACTCC 156
Db 21 IleValLeuLeuIleLeuGlyArgValAsnProValAlaLeuGlyAsnSer-----Pro 38
QY 157 CATAAACCTTATCTCTCACCTGGTTAATTAATCTCACTCCGCCACAGGTTAATATATCAAC 216
Db 39 HisGlnValPheAsnLeuSerTrpGluValThrAsnGluAspArgGlu---ThrValTrp 57
QY 217 AACACTCAAGGGAGGCTCTCTTTAGGAACCTGGTGGCTGATCTATAC-----GTTTGC 270
Db 58 AlaIleThrGlyAsnHisProLeuTrpThrTrpTrpProAspLeuThrProAspLeuCys 77
QY 271 CTCAGATCAGTTATT---CCTAGTCTG-----ACCTCACCC 303
Db 78 MetLeuAlaLeuHisGlyProSerTyrTrpGlyLeuGluTyrGlnAlaProPheSerPro 97
```

QY 304 CCA----- 306  
 Db 98 ProProGlyProProCysSerArgSerGlySerThrProGlyCysSerArgAsp 117  
 QY 306 ----- 306  
 Db 118 CysGluGluProLeuThrSerTyrThrProArgCysAsnThrAlaTrpAsnArgLeuLys 137  
 QY 307 -----GATATCTCTCCATGCTCAC-----GGATTTTATGTTTGCACAGGACCA----- 348  
 Db 138 LeuSerLysValThrHisAlaHisAsnGluGlyPheTyrValCysProGlyProHisArg 157  
 QY 349 CCAATAATGGAACATATGGGAATCCAGAGATTCTTTTGTAAACAATGGAACGT 408  
 Db 158 ProArgTrpAlaArgSerCysGlyGlyProGluSerPheTyrCysAlaSerTrpGlyCys 177  
 QY 409 GTAACCTCTAATGATGGATATTGGAAATGGCCACCTCTCAGCAGGATAGGTAAGTTT 468  
 Db 178 GluThrThrGlyArgAlaSerTrpLys---ProSerSerSerTrpAspTyrIleThr--- 195  
 QY 469 TCTTATGTCAACACCTATPACAGCTCTGTGACAAATTTAATTACCTGACCTGGATTAGAACT 528  
 Db 196 -----ValSerAsnAsnLeuThrSerGlyGln----- 204  
 QY 529 GGAAGCCCCAAGTGCTCTCTTCAGAC---CTAGATTACCTAAATAAAGTTTCACTGAG 585  
 Db 205 AlaThrProValCysLysAsnAsnThrTrpCysAsnSerLeuThrIleArgPheThrSer 224  
 QY 586 AAAGGAAACAAAGAAATATCTTAAATGGCTAAATGGTATGTCTTGGGAAATGGTATAT 645  
 Db 225 LeuGlyLysGlnAlaThr-----SerTrpValThrGlyHisTrpTrpGlyLeuArgLeu 242  
 QY 646 TATGGAGGCTCGGTAAACAAACAGGCTCCATTCTAATTTCCGCTCAAAATAAAGCAG 705  
 Db 243 Tyr---ValSerGlyHisAspProGlyLeuIlePheGlyIleArgLeuLysIleThrAsp 261  
 QY 706 CTGAGGCTCAATGGCTATAGGACCAATAGGTCTTTGAGGGTCAAGACCCCAAC 765  
 Db 262 SerGlyProArgValProIleGlyProAsnProValLeuSerAspGlnArgProProSer 281  
 QY 766 CAAGGACCAGGACCATCTCTTAACATAACTTCTGGATCAGACCCCACTGAGTCTAACAGC 825  
 Db 282 Gln-----ProArgSerProProHisSerAsnSerThrProThrGluThrProLeu 298  
 QY 826 ACG-----ACTAAATGGGGCAAAACTTTTATAGCTTCATCCAGGA 867  
 Db 299 ThrLeuProGluProProAlaGlyValGluAsnArgLeuLeuAsnLeuValLysGly 318  
 QY 868 GCTTTTCAAGCTCTTAATCTCACGACTCCAGAGGCTACCTCTTCTTGTGGCTATGCTTA 927  
 Db 319 AlaTyrGlnAlaLeuAsnLeuThrSerProAspArgThrGlnGluCysTrpLeuLysLeu 338  
 QY 928 GCTTCGGGGCCACCTTACTATGAGGAATGGCTAGAAGAGGAATTCATATGTCACAAA 987  
 Db 339 ValSerGlyProProTyrTyrGluGlyValAlaValLeuGlyThrTyrSerAsnHisThr 358  
 QY 988 GAACATAGAGACCACTGCATCGAGTGGGATCCCAATAAGCTTACCTTACTGAGTTTCT 1047  
 Db 359 SerAlaProAlaAsnCysSerValAlaSerGlnHisLysLeuThrLeuSerGluValThr 378  
 QY 1048 GGAAGAGCCACCTGCATAGGAAGGTTCCCCCAATCCCAACACCTTTGTAAACCACT 1107  
 Db 379 GlyArgGlyLeuCysValGlyAlaValProLysThrHisGlnAlaLeuCysAsnThrThr 398  
 QY 1108 GAAGCCTTTAATCAAACTCTGAGAGTCATATCTGGTACTCTGGTTATGACAGGTGGTGG 1167  
 Db 399 Gln-----AsnThrSerGlyGlySerTyrTyrLeuAlaAlaProAlaGlyThrIleTrp 416  
 QY 1168 GCATGTATATCTGGATTAAACCCCTTGTTGGTTTCCACCTTGGTTTAAACCAAACTAAAGAT 1227  
 Db 417 AlaCysAsnThrGlyLeuThrProCysLeuSerThrThrValLeuAsnLeuThrThrAsp 436  
 QY 1228 TTTTGCATTATGGTCCAAATTTGTTCCCGAGTGATTACTATATCCCGAAAAAGCAATCCTT 1287

Db 437 TyrCysValLeuValGluLeuTrpProArgValThrTyrHisSerProSerTyrValTyr 456  
 QY 1288 GATCAATATGACTACAGAAATCATCGACAAAGAGAGAACCATATCTCTGACACTTCTCT 1347  
 Db 457 HisGlnPheGluGlyArg---AlaLysTyrLysArgGluProValSerLeuThrLeuAla 475  
 QY 1348 GTGATGCTCGGA-----CTTGGAGTGCAGCAGAGGTAGGAAACAGGAAACAGCT 1395  
 Db 476 LeuLeuLeuGlyGlyLeuThrMetGlyGlyIleAlaAlaGlyValGlyThrGlyThrThr 495  
 QY 1396 GCCCTGCTCAGGGACCAACAGCAGCTAGAAACAGAGCTTAGTAACCTACATCGAATTGTA 1455  
 Db 496 AlaLeuVal---AlaThrGlnGlnLeuGlnAlaAla-----Val 507  
 QY 1456 ACAGAAATCTCCAAAGCCCTAGAAAATCTCTCAGTAACCTGGAGGAATCCCTAACCTCC 1515  
 Db 508 HisAspAspLeuLysGluValGluLysSerIleThrAsnLeuGluLysSerLeuThrSer 527  
 QY 1516 TTATCTGAAGTAGTCTTACAGATAGAGAGGTTAGATTATTATTATTTCTAAAAAGAGGA 1575  
 Db 528 LeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeuPheLeuLysGluGly 547  
 QY 1576 GGATATGTGTAGCTTGAAGAGGAATGCTGTTTTTATGTGATCATCTACAGGGCCATC 1635  
 Db 548 GlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrAlaAspHisThrGlyValVal 567  
 QY 1636 AGAGACTCCATGACACAGCTTAGAGAAAGCTTGGAGAGGCTCGAAGGGAAGAAAGAACT 1695  
 Db 568 ArgAspSerMetAlaLysLeuArgGluArgLeuAsnGlnArgGlnLysLeuPheGluSer 587  
 QY 1696 ACTCAAGGGTGTGTTGAGGGATGGTCAACAGAGTCTCTTTGGTTGGCTACCTACTTTCT 1755  
 Db 588 GlyGlnGlyTrpPheGluArgLeuPheAsnGlySerProTrpPheThrThrLeuLysSer 607  
 QY 1756 GCTTTAACAGACACCTTAATAGTCTCTCTCTGTTACTCAGTGTGGCCATGATTATT 1815  
 Db 608 ThrIleMetGlyProLeuIleValLeuLeuLeuIleLeuLeuGlyProCysIleLeu 627  
 QY 1816 AACAAAGTTAATGCTTTCATTAGACAGCAATAGTCAGTCAGTCAGATCATGGTACTTAGA 1875  
 Db 628 AsnArgLeuValGlnPheValLysAspArgIleSerValGlnAlaLeuValLeuThr 647  
 QY 1876 CAACAGTACCAAAAGCCGCTCTAGC 1899  
 Db 648 GlnGlnTyrHisGlnLeuLysSer 655  
 RESULT 6  
 VCMVVR  
 N; polypeptide precursor - radiation murine leukemia virus  
 N; Alternate names: coat polypeptide  
 N; Contains: coat protein gp76; coat protein p15E; R protein  
 C; Species: radiation murine leukemia virus  
 C; Date: 31-Mar-1989 #sequence\_revision 31-Mar-1993 #text\_change 09-Jul-2004  
 C; Accession: C26183  
 R; Merregaert, J.; Janowski, M.; Reddy, E.P.  
 Virol. 158, 88-102, 1987  
 A; Title: Nucleotide sequence of a radiation leukemia virus genome.  
 A; Reference number: A94362; PMID: 87207680; PMID: 3033897  
 A; Accession: C26183  
 A; Molecule type: DNA  
 A; Residues: 1-665 <MER>  
 A; Cross-references: UNIPROT: P11268; UNIPARC: UPI000012A041; GB: M18449; GB: K03363; GB: M18449; NID: G3  
 C; Genetics:  
 A; Gene: env  
 C; Superfamily: type C retrovirus env polypeptide  
 C; Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein  
 F; 1-31/Domain: signal sequence #status predicted <SIG>  
 F; 32-665/Product: env polypeptide #status predicted <ENV>  
 F; 32-604/Domain: extracellular #status predicted <EXT>  
 F; 32-467/Product: coat protein gp76 #status predicted <GP1>  
 F; 468-644/Product: cleavage processing #status predicted <GP2>

F:474-490/Region: hydrophobic #status predicted  
F:605-621/Domain: transmembrane #status predicted <TM1>  
F:622-665/Domain: intracellular #status predicted <INT>  
F:645-665/Product: R protein #status predicted <RP>  
F:43,199,211,324,356,396,400,432/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:  
Pred. No.: 8.67e-99 Length: 665  
Score: 1339.00 Matches: 295  
Percent Similarity: 58.1% Conservative: 93  
Best Local Similarity: 44.2% Mismatches: 180  
Query Match: 38.8% Indels: 100  
DB: 1 Gaps: 21

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x VCMVVR (1-665)

Qy	97	CTTACTCTGTCAATAA	CCTCTCACATAATGGTATGCGCATAGGACAGACGCTGAATCC	156	
		:::	:::		
Db	21	IleValLeuLeuIleLeu	GlyArgValAsnProValAlaLeuGlyAsnSer	Pro 38	
Qy	157	CATAAACCTTATCTCT	CACCTGGCTTAATTAATCACTCCGCGCACAGGTATTAATATCAAC	216	
		:::	:::		
Db	39	HisGlnValPheAsnLeu	SerTrpGluValThrAsnGluAspArgGlu	---ThrValTrp 57	
Qy	217	AACACTCAAGGGAGG	CTCTTTTAGGAACCTGGTGGCTGATCATATAC	---GTTTGC 270	
		:::	:::		
Db	58	AlaIleThrGlyAsnHis	ProLeuTrpThrTrpProAspLeuThrProAspLeuCys	77	
Qy	271	CTCAGATCAGTTAT	---CCTAGTCTG	-----ACCTCAACC 303	
		:::	:::		
Db	78	MetLeuAlaLeuHisGly	ProSerTyTrpGlyLeuGluTyTrpGlnAlaProPheSerPro	97	
Qy	304	CCA	-----	--- 306	
		:::	:::		
Db	98	ProProGlyProProCys	CysSerGlySerSerGlySerThrProGlyCysSerArgAsp	117	
Qy	306	-----	-----	--- 306	
Db	118	CysGluGluProLeuThr	SerTyThrProArgCysAsnThrAlaTrpAsnArgLeuLys	137	
Qy	307	-----GATATCCTCC	ATGCTCAC	-----GGATTTTATGTTGCCAGGACCA	--- 348
		:::	:::		
Db	138	LeuSerLysValThrHis	AlaHisAsnGluGlyPheTyValCysProGlyProHisArg	157	
Qy	349	CCAAATAATGGAACAT	TTCGGGAAATCCAGAGATTTCTTTCTTAAACAATGAACGT	408	
		:::	:::		
Db	158	ProArgTrpAlaArgSer	CysGlyGlyProGluSerPheTyCysAlaSerTrpGlyCys	177	
Qy	409	GTAACCTCTAATGATG	GATATGGAAATGGCCAACTCTCAGCAGGATAGGGTAAGTTT	468	
		:::	:::		
Db	178	GluThrThrGlyArgAla	SerTrpLys---ProSerSerSerTrpAspTyTrpIleThr	--- 195	
Qy	469	TCTTATGTCAACACTTA	TACAGCTCTGGACAATTTAATTACCTGCCTGGATTAGAACT	528	
		:::	:::		
Db	196	-----ValSerAsnAsn	LeuThrSerGlyGln	----- 204	
Qy	529	GGAAGCCCCAAGTGTCT	CTCTTCACAGAC	---CTAGATTACCTAAAAATAAGTTTCACTGAG	585
		:::	:::		
Db	205	AlaThrProValCysLys	LeuAsnThrTrpCysAsnSerLeuThrIleArgPheThrSer	224	
Qy	586	AAAGAAACCAAGAAAA	TATCCTAAAATGGGTAATGGTATGTCTTGGGGAATGGTATAT	645	
		:::	:::		
Db	225	LeuGlyLysGlnAlaThr	-----SerTrpValThrGlyHisTrpTrpGlyLeuArgLeu	242	
Qy	646	TATGAGGCTCGGTAA	CAACACGAGCTCCATCTTAATTCGCTCAAAATAAACGAG	705	
		:::	:::		
Db	243	Tyr---ValSerGlyHis	AspProGlyLeuIlePheGlyIleArgLeuLysIleThrAsp	261	
Qy	706	CTGGAGCTCCAAATG	CGCTTATAGGACCAATACGCTTTCAGCGGTCAAGACCCCAACC	765	
		:::	:::		
Db	262	SerGlyProArgValPro	IleGlyProAsnProValLeuSerAspGlnArgProProSer	281	
Qy	766	CAAGACACGAGGACCAT	CTCTTAAACATACCTTCTGGATCAGACCCCACTGACTCTTAACAGC	825	

Db 628 AsnArgLeuValGlnPheValIysAspArgIleSerValValGlnAlaLeuValLeuThr 647  
 QY 1876 CAACAGTACCAAGCCGCTCTAGC 1899  
 Db 648 GlnGlnTyrHisGlnLeuLysSer 655  
 RESULT 7  
 VCWEEK  
 env polyprotein - AKV murine leukemia virus  
 N:Contains: knob protein gp76; R protein; spike protein p15E  
 C:Species: AKV murine leukemia virus  
 C:Date: 05-Apr-1983 #sequence revision 03-Aug-1984 #text\_change 09-Jul-2004  
 C:Accession: A92995; A93448; A03984  
 R:Herr, W.  
 J. Virol. 49, 471-478, 1984  
 A:Title: Nucleotide sequence of AKV murine leukemia virus.  
 A:Reference number: A92995; MUID:84115072; PMID:6319746  
 A:Accession: A92995  
 A:Molecule type: genomic RNA  
 A:Residues: 1-669 <HER>  
 A:Cross-references: UNIPROT:P03386; UNIPARC:UPI000012A039; GB:J01998; GB:J01999; GB:K000  
 C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein  
 F:1-31/Domain: signal sequence #status predicted <SIG>  
 F:32-470/Product: knob protein gp76 #status predicted <KNS>  
 F:471-650/Product: spike protein p15E #status predicted <SPK>  
 F:651-669/Product: R protein #status predicted <RPT>  
 F:43,199,327,359,399/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 Alignment Scores:  
 Pred. No.: 8.67e-99 Length: 669  
 Score: 1339.00 Matches: 295  
 Percent Similarity: 58.1% Conservative: 94  
 Best Local Similarity: 44.1% Mismatches: 174  
 Query Match: 38.6% Indels: 106  
 DB: 1 Gaps: 22  
 US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x VCWEEK (1-669)  
 QY 97 CTACTCTGTCAATPAACCTCTCAGACTAATGGTATGGCATAGGACAGCAGCTGAATCC 156  
 Db 21 IleValLeuLeuLeuLeuGlyGlyValAsnProValThrLeuGlyAsnSer-----Pro 38  
 QY 157 CATAAACCTTATCTCTACCTGGTAAATTAATGACTCCGGCAGAGTATTAATATCAAC 216  
 Db 39 HisGlnValPheAsnLeuThrTrpGluValThr-----AsnGlyAspArgGluThrValTrp 57  
 QY 217 AACACTCAAGGGAGGCTCTTTAGGAACCTGGTGGCTGATCTATAC-----GTTTGC 270  
 Db 58 AlaIleThrGlyAsnHisProLeuThrTrpTrpProAspLeuThrProAspLeuCys 77  
 QY 271 CTCAGATCAGTAFTT---CTAGTCTG-----ACCTCACCC 303  
 Db 78 MetLeuAlaLeuHisGlyProSerTyrTrpGlyLeuGluTyrArgAlaProPheSerPro 97  
 QY 304 CCA----- 306  
 Db 98 ProGlyProProCysCysSerGlySerSerAspSerThrProGlyCysSerArgAsp 117  
 QY 306 ----- 306  
 Db 118 CysGluGluProLeuThrSerTyrThrProArgCysAsnThrAlaTrpAsnArgLeuLys 137

QY 307 -----GATATCCTCCATGCTCAC-----GGATTTTATGTTGTCACAGGACCA----- 348  
 Db 138 LeuSerLysValThrHisAlaHisAsnGlyGlyPheTyrValCysProGlyProHisArg 157  
 QY 349 CCAATAATGGAACAACATTCGGGAATCCAGAGATTCTTTTGTAAACAATGGAACTGT 408  
 Db 158 ProArgTrpAlaArgSerCysGlyGlyProGluSerPheTyrCysAlaSerTrpGlyCys 177  
 QY 409 GTAACCTCTAATGATGATATTGAAATGGCAACCTCTCAGCAGGATGAGGTAGTTT 468  
 Db 178 GluThrThrGlyArgAlaSerTrpLys-----ProSerSerTrpAspTyrIleThrVal 196  
 QY 469 TCTTATGTCAACACTATACAGCTCTGGACAATTTAATTACCTGACCTGGATTAGAACT 528  
 Db 197 Ser-----AsnAsnLeuThrSerAsp-----Gln 204  
 QY 529 GGAAGCCCAAGTCTCTCTTCAGAC-----CTAGATTACCTAAATAAAGTTTCACTGAG 585  
 Db 205 AlaThrProValCysLysGlyAsnGluTrpCysAsnSerLeuThrIleArgPheThrSer 224  
 QY 586 AAGGAAAAACAAGAAAAATATCTAAAAATGGTAAATGGTATGTTCTGGGGAATGTATAT 645  
 Db 225 PheGlyLysGlnAlaThr-----SerTrpValThrGlyHisTrpTrpGlyLeuArgLeu 242  
 QY 646 TAGGAGGCTCGGTAAACAACAGGCTCCATTAATACTATTCGCCTCAAAATAAACCCAG 705  
 Db 243 Tyr---ValSerGlyHisAspProGlyLeuIlePheGlyIleArgLeuLysIleThrAsp 261  
 QY 706 CTGGAGCTCCATGGCTATAGACCAAAATACGGTCTTTCGCGGTCAAGACCC----- 759  
 Db 262 SerGlyProArgValProIleGlyProAsnProValLeuSerAspArgProProSer 281  
 QY 760 -----CAACCCCAAGCAGGACCACTCTCTAACATACTCTGGATCAGACCCC 810  
 Db 282 ArgProArgProThrArgSerProProSerAsn-----SerThrPro 296  
 QY 811 ACTGAGTCTAACAGCAGC-----ACTAAATGGGGGCCAAACTTTT 852  
 Db 297 ThrGluThrProLeuThrLeuProGluProProAlaGlyValGluAsnArgLeuLeu 316  
 QY 853 AGCTCATCCAGGAGGCTTTTCAAGCTCTTAATCCAGACTCCAGAGGCTACCTCTTCT 912  
 Db 317 AsnLeuValLysGlyAlaTyrGlnAlaLeuAsnLeuThrSerProAspLysThrGlnGlu 336  
 QY 913 TGTGGCTATGCTTAGCTTCGGGCCACCTTACTATGAAGGAATGGCTAGAGAGGAAA 972  
 Db 337 CysTrpLeuCysLeuValSerGlyProProTyrTrpGlyValAlaValLeuGlyThr 356  
 QY 973 TTCATGTGCAAAAGAACAATAGAGACCAATGCACATGGGGATCCCAAAATAAGCTTACC 1032  
 Db 357 TyrSerAsnHisThrSerAlaProAlaAsnCysSerValAlaSerGlnHisLysLeuThr 376  
 QY 1033 CTTACTGAGGTTTCTGGAAGGACCTGATAGGAAGGTTCCTCCCATCCCAACACAC 1092  
 Db 377 LeuSerGluValThrGlyGlnGlyLeuCysIleGlyAlaValProLysThrHisGlnVal 396  
 QY 1093 CTTTGTAAACACACATGAGGCTTTAATCAACCTCTGAG---AGTCAATATCTGGTACCT 1149  
 Db 397 LeuCysAsnThrThrGln-----LysThrSerAspGlySerTyrTyrLeuAlaAla 413  
 QY 1150 GGTATGACAGGTTGGGGCATGTAATCTATCTGGATTAACCCCTTGTGTTCCACCTGGTT 1209  
 Db 414 ProThrGlyThrThrTrpAlaCysSerThrGlyLeuThrProCysIleSerThrThrIle 433  
 QY 1210 TTTAACCAACTAAAGATTTTTCATATGATGGTCCAAATTTGTTCCCGAGGTATTACTAT 1269  
 Db 434 LeuAspLeuThrThrAspTyrCysValLeuValGluLeuTrpProArgValThrThrHis 453  
 QY 1270 CCGGAAAAACCAATCCTTGATGATATGACTACAGAAATCATCGACAAAAGAGACCCC 1329  
 Db 454 SerProSerTyrValTyrHisGlnPheGlu---ArgArgAlaLysTyrLysArgGluPro 472  
 QY 1330 ATATCTCTGACACTTGTGTGATGCTCGGA-----CTTGAGTGGCAGCAGGT 1377









Db 223 ThrProAlaCysLysGlyAsnLysTrpCysAsnSerLeuThrIleArgPheThrSerPhe 242  
QY 589 GGAACAACAGAAATATCTAAATGGTAAATGGTATGCTCTGGGGAATGTATATTAT 648  
Db 243 GlyLysGlnAlaThr-----SerTrpValThrGlyHisTrpTrpGlyLeuArgLeuTyr 260  
QY 649 GGAGGCTGGGTAAACACAGGCTCCATCTTAACCTATTCGCTCAAAATAAACAGCTG 708  
Db 261 ---ValSerGlyHisAspProGlyLeuIlePheGlyLeuArgLeuLysIleThrAspLeu 279  
QY 709 GAGCTCCATCGCTATAGGACCAATACGGTCTTACGGGTCAAGACCC-----759  
Db 280 GlyProArgValProIleGlyProAsnProValLeuSerAspArgArgProSerArg 299  
QY 760 -----CCAACCCAGGACCGACCATCTCTAATACTTCGGATCAGACCCCACT 813  
Db 300 ProArgProThrArgSerProProSerAsn-----SerThrProThr 314  
QY 814 GAGTCTACAGCAG-----ACTAAATGGGGGCAAACTTTTATAGC 855  
Db 315 GluThrProLeuThrLeuProGluProProAlaGlyValGluAsnArgLeuLeuAsn 334  
QY 856 CTCATCCAGGAGCTTTTCAAGCTCTTAAGCTCCAGCTCCAGAGGCTACCTCTTCTGT 915  
Db 335 LeuValLysGlyAlaTyrGlnAlaLeuAsnLeuThrSerProAspLysThrGlnGluCys 354  
QY 916 TGGCTATCTTAGCTTCGGGCCACCTTACTATGAAGGAATGGCTAGAGAGGGAATTC 975  
Db 355 TrpLeuCysLeuValSerGlyProProTyrTrpGluGlyValAlaValLeuGlyThrTyr 374  
QY 976 AATGTGACAAAGAACATAGAGACCAATGCATCGGATCCCAATAAAGCTTACCTT 1035  
Db 375 SerAsnHisThrSerAlaProAlaAsnCysSerValAlaSerGlnHisLysLeuThrLeu 394  
QY 1036 ACTGAGGTTCTCGAAAGGACCTGCATAGCAAGGTTCCCATCCACCAACACCTT 1095  
Db 395 SerGluValThrGlyGlnGlyLeuCysIleGlyAlaValProLysThrLeuGlnValLeu 414  
QY 1096 TGTAAACCACTGAAGCCCTTAAATCAACCTCTGAGAGTCAATATCTGCTGGTGGTTAT 1155  
Db 415 CysAsnThrThrGln-----LysThrSerAlaGlySerTyrTyrLeuAlaAlaProThr 432  
QY 1156 GACAGGTGGTGGCATGTAATCTGATTAACCCCTTGTGTTTCCACCTGGTGGTTTAAAC 1215  
Db 433 GlyThrIleTrpAlaCysSerThrGlyLeuThrProCysIleSerThrIleLeuAsn 452  
QY 1216 CAATTAAGATTTTTCATTTATGTCCTCAAAATGTTCCCGAGTGTATCTATCCCGAA 1275  
Db 453 LeuThrThrAspTyrCysValLeuValGluLeuTrpProArgValThrTyrHisSerPro 472  
QY 1276 AAAGCAATCTTGATGAATATGACTACAGAAATCATCGACAAAGAGAGAACCCATATCT 1335  
Db 473 SerTyrValTyrHisGlnPheGlu---ArgArgAlaLysTyrLysArgGluProValSer 491  
QY 1336 CTGACACTTGTGTGATGCTCGGA-----CTTGAGTGGCAGCAGGTGTAGGA 1383  
Db 492 LeuThrLeuAlaLeuLeuLeuGlyGlyLeuThrMetGlyGlyIleAlaAlaGlyValGly 511  
QY 1384 ACAGGACAGCTGCCCTGGTCACGGGACCAAGCAGCTAGAAACAGGACTAGTAACCTA 1443  
Db 512 ThrGlyThrThrAlaLeuValAlaThrGlnGln-----PheGlnGlnLeu 526  
QY 1444 CATCGAATTTGAACAGAGATCTCCAAGCCCTAGAAAATCTGTCAAGTAACCTGAGGAA 1503  
Db 527 GlnAlaAlaMetHisAspAspLeuLysGluValGluLysSerIleThrAsnLeuGluLys 546  
QY 1504 TCCCTAACCTCTTATCTGAAGTATGCTTACAGAAATGAAGAGGTTAGATTATTATTT 1563  
Db 547 SerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuPhe 566  
QY 1564 CTAAGACAGGAGATTTATGTTAGCTTGNAGGAGGAATGCTCTTTTATGTCGATCAT 1623  
Db 567 LeuLysGluGlyGlyLeuCysAlaAlaLeuLysGluCysPheTyrAlaAspHis 586

QY 1624 TCAGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGCTTGAGAAAGCTCGAAGG 1683  
Db 587 ThrGlyLeuValArgAspSerMetAlaLysLeuArgGluArgLeuSerGlnArgGlnLys 606  
QY 1684 GAAAGGAAACTACTCAAGGGTGGTTGAGGATGGTTCAACAGGTCTCTTTGGTGGCT 1743  
Db 607 LeuPheGluSerGlnGlnGlyTrpPheGluGlyLeuPheAsnLysSerProTrpPheThr 626  
QY 1744 ACCCTACTTCTGCTTTAAACAGGACCTTAATAGTCCTCTCTCTTACTACACTTGGG 1803  
Db 627 ThrLeuIleSerThrIleMetGlyProLeuIleIleLeuLeuLeuLeuPheGly 646  
QY 1804 CCATGTATTATTAAACAGTTAAATTCCTTCAATTAGAGAACGAATAAGTCAGTCAGATC 1863  
Db 647 ProCysIleLeuAsnArgLeuValGlnPheIleLysAspArgIleSerValValGlnAla 666  
QY 1864 ATGCTACTTAGACACAGTACCAAGCCCGTCTAGC 1899  
Db 667 LeuValLeuThrGlnGlnTyrHisGlnLeuLysSer 678  
RESULT 11  
VCVWEM  
env polyprotein - Moloney murine leukemia virus  
N:Contains: knob protein gp70; R protein; spike protein p15e  
C:Species: Moloney murine leukemia virus  
C>Date: 01-Sep-1981 #sequence revision 24-Sep-1981 #text\_change 09-Jul-2004  
C:Accession: A93265; A93235; A93848; A03983  
R:Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.  
Nature 293, 543-548, 1981  
A:Title: Nucleotide sequence of Moloney murine leukaemia virus.  
A:Reference number: A93265; MUID:82035843; PMID:6169994  
A:Accession: A93265  
A:Molecule type: genomic RNA  
A:Residues: 1-565 <SH>  
A:Cross-references: UNIPROT:P03385; UNIPARC:UPI00000134DC; GB:J02255; GB:J02256; GB:J0.  
A:Experimental source: clone pMLV-1  
R:Sutcliffe, J.G.; Shinnick, T.M.; Green, N.; Liu, P.T.; Niman, H.L.; Lerner, R.A.  
Nature 287, 801-805, 1980  
A:Title: Chemical synthesis of a polypeptide predicted from nucleotide sequence allows  
A:Reference number: A93235; MUID:81052384; PMID:6159543  
A:Accession: A93235  
A:Molecule type: genomic RNA  
A:Residues: 496-665 <SUT1>  
A:Cross-references: UNIPARC:UPI0000174A47  
A:Experimental source: provirus  
R:Sutcliffe, J.G.; Shinnick, T.M.; Verma, I.M.; Lerner, R.A.  
Proc. Natl. Acad. Sci. U.S.A. 77, 3302-3306, 1980  
A:Title: Nucleotide sequence of Moloney leukemia virus: 3' end reveals details of repli  
A:Reference number: A93848; MUID:81013872; PMID:6251454  
A:Accession: A93848  
A:Molecule type: DNA  
A:Residues: 484-662, 'CEF' <SUT2>  
A:Cross-references: UNIPARC:UPI0000174A48  
A:Experimental source: provirus, clone pMLV-201  
C:Genetics:  
A:Gene: env  
C:Superfamily: type C retrovirus env polyprotein  
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane prote  
F:34-469/Product: knob protein gp70 #status predicted <KNB>  
F:470-649/Product: spike protein p15e #status predicted <SPK>  
F:650-665/Product: R protein #status predicted <RPT>  
F:45,199,356,358,398,434/Binding site: carbohydrate (Asn) (covalent) #status predicted  
Alignment Scores:  
Pred. No.: 3,43e-97 Length: 665  
Score: 1319.00 Matches: 291  
Percent Similarity: 55.7% Conservative: 86  
Best Local Similarity: 43.0% Mismatches: 198  
Query Match: 38.0% Indels: 102  
DB: 1 Gaps: 20  
US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x VCVWEM (1-665)



Qy	49	CCGAAAGACTGAAATCCCCTTAAGCTTCGGCTCCATCGCTGGTTCCTTACTCTGTCA	108
Db	16	ProArgGlyProLeuLeileProLeuLeuMetLeuArgGlyValSerThrAlaSer	35
Qy	109	ATAACCTCTCAGACTAATGGTATGCGCATAGAGACAGCCTCACTCCCATAAACCCTTA	168
Db	36	ProGlySer-----SerProHisGlnValTyr	44
Qy	169	TCTCTCACTGGTTAATTACTGACTCCGGCAGAGGTATTAATATCAACAACACTCAAGG	228
Db	45	AsnIleThrTrpGluValThr---AsnGlyAspArgGluThrValTrpAlaThrSerGly	63
Qy	229	GAGGCTCCTTTAGGAACCTGGGCGCTGATCATAC-----GTTTGCCTC-----	273
Db	64	AsnHisProLeuTrpThrTrpProAspLeuThrProAspLeuCysMetLeuAlaHis	83
Qy	274	-----AGATCAGTTATTTCCTAGTCTGACCTCACCC	303
Db	84	HisGlyProSerTyrTrpGlyLeuGluTyrGlnSerProPheSerSerProGlyPro	103
Qy	304	CCA-----	306
Db	104	ProCysCysSerGlyGlySerSerProGlyCysSerArgAspCysGluGluProLeuThr	123
Qy	307	-----GATATCCTCCAT	318
Db	124	SerLeuThrProArgCysAsnThrAlaTrpAsnArgLeuLysLeuAspGlnThrThrHis	143
Qy	319	GCT-----CACGGATTTATGTTTCCCGACGACCA-----CCAATATCGGAACAT	366
Db	144	LysSerAsnGluGlyPheTyrValCysProGlyProHisArgProArgGluSerLysSer	163
Qy	367	TGCGGAAATCCAGAGATTTCTTTGTAAACAATGGAACCTGTGTAACTCTAATGATGCA	426
Db	164	CysGlyGlyProAspSerPheTyrCysAlaTyrTrpGlyCysGluThrThrGlyArgAla	183
Qy	427	TATTGGAAATGCCAACCTCTCAGCAGGATAGGGTAAGTTTTTCTTATGTCACACCTAT	486
Db	184	TyrTrpLys---ProSerSerSerTrpAsp-----PheIleThrValAsnAsnAsn	199
Qy	487	ACCAGCTCTGGACAATTAATTACCTGCACCTGGATTAGAACTGGAACTCCCAAGTCTCT	546
Db	200	LeuThrSerAspGln-----AlaValGlnValCysLysAspAsnLysTrpCysAsn	216
Qy	547	CCTTCAGACCTAGATTACCTAAAAATAAGTTTTCACGTAGAGAAAGAAAACAAGAAATATC	606
Db	217	Pro-----LeuValIleArgPheThrAspAlaGlyArg-----Val	229
Qy	607	CTAAATGGTAAATGGTATGCTTGGGGAATGGTATATTATGAGGCTCGGGTAAACAA	666
Db	230	ThrSerTrpThrThrGlyHisTyrTrpGlyLeuArgLeuTyr---ValSerGlyGlnAsp	248
Qy	667	CCAGGCTCCATTCTTAACATTTCGCTCAAAATAAACACAGCTGAGGCTCCAAATGGCTATA	726
Db	249	ProGlyLeuThrPheCylleargLeuArgTyrGlnAsnLeuGlyProArgValProIle	268
Qy	727	GGACCAATACGGTCTTTCGCGGTCAAGACCC-----CCAAACCAAGSACCAGGA	777
Db	269	GlyProAsnProValLeuAlaAspGlnGlnProLeuSerLysPProLysProValLysSer	288
Qy	778	CCATCCTCTAATCACTTCTCGATCAGACCCCACTGAGTCTTAACAGCAGGACTAAATG	837
Db	289	ProSerValThrLysProProSerGlyThrProLeuSerProThrGlnLeuProProAla	308
Qy	838	GGGGCA-----AAACTTTTACGCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCACG	891
Db	309	GlyThrGluAsnArgLeuLeuAsnLeuValAspGlyAlaTyrGlnAlaLeuAsnLeuThr	328
Qy	892	ACTCCAGAGGCTACCTCTCTTTGTGGCTATGCTTAGCTTCGGGCCCACTTACTATGAA	951
Db	329	SerProAspLysThrGlnGluCysTrpLeuCysLeuValAlaGlyProTyrTrpGlu	348

QY	952	GGAAATGCTAGAGAGGGAAATTCATATGTGACAAAAGACATAGAGACCATCCATCGG	1011
DB	349	GlyValAlaValLeuGlyThrTyrSerAsnHisThrSerAlaProAlaAsnCysVal	368
QY	1012	GGATCCCAAAATAAGCTTACCTTACTGTAGGTTTTCGGAAAGGACCTGCATAGGAAG	1071
DB	369	AlaserGlnHisIysLeuThrLeuSerGluValThrGlyGlnGlyLeuCysIleGlyAla	388
QY	1072	GTTCCTCCCATCCCAACCAACCTTTGTAAACACACATGAAGCCTTTAATCAAACTCTGAG	1131
DB	389	ValProIysThrHisGlnAlaLeuCysAsnThrThr-----GlnThrSerSer	404
QY	1132	-----AGTCAATATCTGGTACCTGGTTATGACAGAGTGGGCATGTAATACTGATTA	1185
DB	405	ArgGlySerTyrTyrLeuValAlaProThrGlyThrMetTrpAlaCysSerThrGlyLeu	424
QY	1186	ACCCCTGTGTTCCCACTTGGTTTAAACAACTAAAGATTTTGTGCATTATGTGTCCAA	1245
DB	425	ThrProCysIleSerThrThrIleLeuAsnLeuThrThrAspTyrCysValLeuValGlu	444
QY	1246	ATTGTTCCCGAGTGTTACTATCCCGAAAAAGCAATCCTTGATGAATATGACTACAGA	1305
DB	445	LeuTrpProArgValThrTyrHisSerProSerTyrValTyrGlyLeuPheGlu--Arg	463
QY	1306	AATCATCGACAAAGAGAGAACCCATATCTCTGACACTGTGCTGTGATGTCGGA-----	1359
DB	464	SerAsnArgHisIysArgGluProValSerLeuLeuAlaLeuLeuGlyLeuGlyLeu	483
QY	1360	-----CTTGGATGGCAGCAGCTGTAGGAACAGGACAGCTCCCTGGTCACGGGACCA	1413
DB	484	ThrMetGlyGlyIleAlaGlyIleGlyThrGlyThrThrAlaLeuMetAlaThrGln	503
QY	1414	CAGCAGCTAGAAACAGGACTTAGTAACATCGAATTGTAAACAGAGAATCTCCAGGCC	1473
DB	504	Gln-----PheGlnGlnLeuGlnAlaValGlnAspAspLeuArgGlu	518
QY	1474	CTAGAAAAATCTGTCTAGTAACCTGGAGGAATCTCTAACCTCTTATCTGAAGTAGTCCTTA	1533
DB	519	ValGluIysSerIleSerAsnLeuGluLysSerLeuThrSerLeuSerGluValValLeu	538
QY	1534	CAGAAATAGAGGGTTAGATTATATTATTCTAAAGAGGAGATTATGTGTAGCCTTG	1593
DB	539	GlnAsnArgArgGlyLeuAspLeuLeuPheLeuLysGluGlyLeuCysAlaIleLeu	558
QY	1594	AAGAGAGAAATCTGTTTTTATGTGGATCATTTACGGGGCCATCAGAGACTCCATGAACAAG	1653
DB	559	LysGluGluCysCysPheTyrAlaAspHisThrGlyLeuValArgAspSerMetAlaLys	578
QY	1654	CTTAGAGAAAGGTTGGAGAAGCGTCGAGGGGAAAAGGAACCTACTCAGGGCTGGTTGAG	1713
DB	579	LeuArgGluArgLeuAsnGlnArgGlnLysLeuPheGluSerThrGlnGlyTrpPheGlu	598
QY	1714	GGATGGTTCAACAGTCTCTTTGGTTGGCTTACCTACTTTCTGCTTTTAAACAGACCCCTTA	1773
DB	599	GlyLeuPheAsnArgSerProTrpPheThrThrLeuIleSerThrIleMetGlyProLeu	618
QY	1774	ATAGTCTCTCTCTGTACTCACAGTTGGGCCCATGTATTATTAAACAGTTTAATTCGCTTC	1833
DB	619	IleValLeuLeuMetIleLeuLeuPheGlyProCysIleLeuAsnArgLeuValGlnPhe	638
QY	1834	ATTAGAGAACGAATAAGTCGAGTCCAGATCATCGTACTTAGACAACACAGTAC	1884
DB	639	ValIysAspArgIleSerValValGlnAlaLeuValLeuThrGlnGlnTyr	655

RESULT 12

VCMVHL

env polyprotein precursor - HoMuLV murine leukemia virus

N: Contains: knob protein gp76; spike protein p15E  
env polyprotein precursor - homologous murine leukemia

N; Contains: knob protein gp6; spike protein gp7; leukemia virus

C;Species: *HoMuLV murine leukemia virus*  
A:Note: host *Mus hortulanus* (European mouse)

A;Note: host Mus hortulanus (European mouse)  
C;Date: 31-Dec-1990 #sequence revision 31-Dec-1990 #text change 09-Jul-2004

C;Date: 31-Dec-1990  
C:Accession: B32594

C;Accession: B32594  
R:Vovtek. P.: Kozak. C.A.

Virology 173, 58-67, 1989  
 A:Title: Nucleotide sequence and mode of transmission of the wild mouse ecotropic virus,  
 A:Reference number: A32594; MUID:90051094; PMID:2554579  
 A:Accession: B32594  
 A:Molecule type: DNA  
 A:Residues: 1-666 <VOY>  
 A:Cross-references: UNIPROT:P21436; UNIPARC:UPI000012A03F  
 A:Note: the authors translated the codon CTT for residue 451 as Pro  
 C:Genetics:  
 A:Gene: env  
 C:Superfamily: type C retrovirus env polyprotein  
 C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein  
 F:1-30/Domain: signal sequence #status predicted <SIG>  
 F:31-466/Product: knob protein gp76 #status predicted <KPG>  
 F:467-666/Product: spike protein p15E #status predicted <SPP>  
 F:473-489/Domain: transmembrane #status predicted <TM1>  
 F:607-623/Domain: transmembrane #status predicted <TM2>  
 F:42,197,290,324,356,363,431,599/Binding site: carbohydrate (Asn) (covalent) #status pre

## Alignment Scores:

Pred. No.:	1.03e-96	Length:	666
Score:	1313.00	Matches:	300
Percent Similarity:	57.4%	Conservative:	95
Best Local Similarity:	43.6%	Mismatches:	182
Query Match:	37.9%	Indels:	111
DB:	1	Gaps:	23

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x VCMVHL (1-666)

Qy	97	CTTACTCTGTCATTAACCTCTCAGACTAATGGT-----	129
Db	6	LeuProLysSerHisLeuValThrAsnProTrpGlyProIleLeuLeuGlyLeu 25	
Qy	130	ATGCGCATAGGAGACAGCTG-----AACTCCCAATAAACCTTATCTCAGCTGG 180	
Db	26	IleMetLeuGlyAlaLeuGlySerProHisLeuValPheAsnLeuThr 45	
Qy	181	TTAATTACTGATCCGGCACAGGTATTAAATACACACTCAAGGGAGCTCTTTA 240	
Db	46	GluValTyrAsnGlnGluTyrGlu---ThrValTrpAlaThrSerGlySerHisProLeu 64	
Qy	241	GGAACTGGTGGCTGATCTATAC-----GTTGGCTCAGATCAGTATT----- 285	
Db	65	TrpThrTrpTrpProThrLeuProAspLeuCysMetLeuAlaGlnLeuAlaLysPro 84	
Qy	286	-----CCTAGTCTGACCTCACCCAGAT----- 309	
Db	85	SerTrpGlyLeuSerAspTyrProProTyrSerLysProGlyProProCysCysThr 104	
Qy	309	----- 309	
Db	105	ThrAspAsnAsnProProGlyCysSerArgAspCysAsnGlyProLeuThrTyrLeuThr 124	
Qy	310	-----ATCTCCATGCTCAC----- 324	
Db	125	ProArgCysSerThrAlaTrpAsnArgLeuLysLeuValLeuThrThrHisLeuAsn 144	
Qy	325	---GGATTATGTTTCCAGGACCA-----CCAAATAATGAAACATTCGGGAAT 375	
Db	145	GlnGlyPheTyrValCysProGlyProHisArgProArgHisAlaArgAsnCysGlyGly 164	
Qy	376	CCAGAGATTCTTTTGTAAACATGGAACCTGTAACTCTTAATGATGATATGGAAA 435	
Db	165	ProAspAspPheTyrCysAlaHisTyrGlyCysGluThrThrGlyGlnAlaTyrTyrLys 184	
Qy	436	TGGCCAACTCTCAGCAGGATAGGTAAATTTTCTTATGTCACACTATACAGCTCT 495	
Db	185	---ProSerSerTrpAspTyrIleArgValSer-----AsnAsnAlaSerSerSer 201	
Qy	496	GGCAATTTAATTACTGACTGATAGTAACTGGAAGCCCAAGCTCTCTCTCAGAC 555	
Db	202	Asp-----AlaThrAlaCysLysAsnAsnAsnTrpCysSerPro----- 215	

Qy	556	CTAGATTACCTAAATAAAGTTTCACTGAGAAAGAAACAAGAAAAATATCTTAAATGG 615
Db	216	-----LeuAlaIleSerPheThrAspProGlyLysArgAlaThr-----SerTrp 230
Qy	616	GTAATGGTATCTCTGGGAATG---GTATATTATGGAGGCTCGGGTAAACAACAGGC 672
Db	231	ThrSerGlyPheThrTrpGlyLeuArgLeuTyrIleSerGly-----HisProGly 247
Qy	673	TCCATTCTAATATTTCGCTCAAATAAACACAGCTGGAGCTCCCAATGGCTATAGGACA 732
Db	248	LeuIlePheGlyValArgLeuLysIleSerAspLeuGlyProArgValProIleGlyPro 267
Qy	733	AATACGGCTTTCAGCGGTCAAAGACCCCAACCAAGGACCA-----GGACCA 780
Db	268	AsnProValLeuSerGluGlnArgProProSerGlnProGluProAlaArgLeuProPro 287
Qy	781	TCCTCTAACATAACTCTTCGATCAGACCCCACTGAGTCTAACAGCAGCACTAAATGGG 840
Db	288	SerSerAsnLeuThrGlnGlyGlyThrProSerAlaProThrGlyProProGlnGly 307
Qy	841	GCA-----AAACTTTTTCAGCTCATCCAGGAGCTTTCAAGCTCTTAACTCCACGACT 894
Db	308	ThrGlyAspArgLeuLeuAspLeuValGlnGlyAlaTyrGlnAlaLeuAsnAlaThrSer 327
Qy	895	CCAGAGCTACCTCTCTTGTGGCTATGCTTTCGGGCCCACTTACTATGAAGGA 954
Db	328	ProAspLysThrGlnGluCysTrpLeuCysLeuValSerSerProProTyrTyrGlyGly 347
Qy	955	ATGCTAGAGAGGGAATTCATGTGACAAAGAACATGAGACCAATGCACATGGGGA 1014
Db	348	ValAlaValValGlyProTyrSerAsnHisThrAlaProAlaAsnCysSerAlaAsp 367
Qy	1015	TCCCAATAAAGCTTACCTTACTGAGGTTCTCGAAAGGCACCTGCATAGGAAGGT 1074
Db	368	SerGlnHisLysLeuThrLeuSerGluValThrGlyLys---ProLeuProArgLysGly 387
Qy	1075	CCCCATCCCAACCAACCTTTGTAAACACACATGAAGCTTTAATCAACCTCTGAGAGT 1134
Db	387	erGlnAspProProGlyProValGlnTyrHis-----SerGly-AlaArgGln 402
Qy	1135	CAATATCTGGTACCTGGTTATGACAGG-----TGGTGGGATGTATATCTGGATTAC 1189
Db	403	LysTyrSerLeuSerGlyGlySerArgGlyThrMetTrpAlaCysAsnThrGlyLeuThr 422
Qy	1189	CTTGTCTTCCCACTTGGTTTTTAAACCAACATAAGATTTTCATTATGCTCAAT 1248
Db	423	ProCysLeuSerThrAlaValLeuAsnLeuThrThrAspTyrCysValLeuValGluLeu 442
Qy	1249	GTTCCTCCGAGTGATTACTATATCCGAAAAAGCAATCTTGTATGAATATGACTACAGAA 1308
Db	443	TrpProArgValThrTyrHisSerLeuAspPheValTyrArgGlnValGlyArgThr 462
Qy	1309	CATCGACAAAGAGAGAACCATATCTCTGACACTTCTGCTGTGATGCTCGGA----- 1359
Db	463	---ArgTyrGlnArgGluProValSerLeuThrLeuAlaLeuLeuGlyGlyLeuThr 481
Qy	1360	---CTTCGAGTGCAGCAGGTGTAGGAACAGGACAGCTCCCTCGTGCACGGACCA 1416
Db	482	MetGlyGlyIleAlaAlaGlyValGlyThrGlyThrSerAlaLeuValLys---ThrGln 500
Qy	1417	CAGCTAGAACACAGGACTTAGTAACTACATCGAATTTGTAAACAGAAAGATCTCCAGCCCTA 1476
Db	501	GlnPheGlu-----GlnLeuHisAlaAlaIleGlnAlaAspLeuLysGluVal 516
Qy	1477	GAATAATCTCTAGTAACCTGGAGGAATCCCTAACCTCTTATCTGAAGTAGTCTCAG 1536
Db	517	GluSerSerIleThrAsnLeuGluLysSerLeuThrSerLeuSerGluValValLeuGln 536
Qy	1537	AATAGAGAGGGTTAGATTTTATTATTAAGAGGAGGATTTATGTAGCTCTCAAG 1596
Db	537	AsnArgGlyGlyLeuAspLeuPheLeuGlyGlyGlyLeuCysAlaLeuLys 556
Qy	1597	GAGGAATGCTGTTTTTTATGTGGATCATTCAGGGGCCATCAGAGACTCCATGAACAGCTT 1656

```
Db 557 GluGluCysCysPheTyrAlaAspHisThrGlyLeuValArgAspSerMetAlaLysLeu 576
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1657 AGAGAAAGGTTGGAGAACGGTCGAGGAGGAAAGGAACTACTCAAGGTTGGTTGAGGGA 1716
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 577 ArgGluArgLeuAsnGlnArgGlnLysLeuPheGluAlaGlyGlnGlyTrpPheGluGly 596
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1717 TGGTTCAACAGCTCTTTGGTTGGCTACCTACTTCTGCTTTAAACAGGACCTTAATA 1776
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 597 LeuPheAsnArgSerProTrpPheThrLeuIleSerThrIleMetGlyProLeuLeu 616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1777 GTCCTCTCTCTTACTCACAGTTGGGCATGTATTTAAACAAGTTAAATGCGCTTCATT 1836
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 617 IleLeuLeuLeuIleLeuMetPheGlyProCysIleLeuAsnArgLeuValGlnPheVal 636
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1837 AGAGAACGAATAAGTCAGTCAGATCATGTTAGTACTTAGACACAGTACCAAGCCCGTCT 1896
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 637 LysAspArgIleSerValValGlnAlaLeuValLeuThrGlnGlnTyrHisGlnLeuLys 656
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1897 AGCAGGGAAGCTGGCCGC 1914
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 657 ProLeuGluHisGlyArg 662
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
VCMVRV
env polyprotein precursor - Rauscher mink cell focus-forming virus
N:Contains: coat protein p12E; coat protein p15E; knob protein gp70
C:Species: Rauscher mink cell focus-forming virus
C>Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 09-Jul-2004
C:Accession: A03990
R:Vogt, M.; Haggbloom, C.; Swift, S.; Haas, M.
J. Virol. 55, 184-192, 1985
A:Title: Envelope gene and long terminal repeat determine the different biological properties of Rauscher mink cell focus-forming virus
A:Reference number: A93011; PMID:85237696; PMID:4009793
A:Accession: A03990
A:Molecule type: DNA
A:Residues: 1-640 <VOG>
A:Cross-references: UNIPROT:P06445; UNIPARC:UPI000012A047; GB:M10100; NID:9332068; PIDN:
C:Gene: env
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F:1-32/Domain: signal sequence #70 status predicted <SIG>
F:33-443/Product: knob protein gp70 #status predicted <KBP>
F:444-640/Product: coat protein p15E #status predicted <PFB>
F:444-623/Product: coat protein p12E #status predicted <PTB>
F:443,58,300,332,339,372,408,576/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 7,066-95 Length: 640
Score: 1290.00 Matches: 277
Percent Similarity: 59.9% Conservative: 104
Best Local Similarity: 43.6% Mismatches: 171
Query Match: 37.2% Indels: 84
DB: 1 Gaps: 19

US-10-723-552-3_COPY_5620_7533 (1-1914) x VCMVRV (1-640)
Qy 130 ATGCGCATAGGACAGACGCTG-----AACTCCCAATAAACCTTATCTCTCACTGG 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 27 IleArgAlaGlyValSerValGlnHisAspSerProHisLysValPheAsnValThrTrp 46
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 181 TTAATTACTAGCTCCGGCAGGATTAATATACACACACTCAAGGAGGAGCTCTCTTA 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 47 ArgValThrAsnLeuMetThrGlyGlnThrAlaAsnAlaThr-----SerLeuLeu 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 241 GGAACCTCG-----TGGCTCATCTATACGTTTGGCTCAGATCAGTTATCTTAGT 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 GlyThrMetThrAspAlaPheProLysLeuTyrPheAspLeuCysAspLeuValGlyAsp 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy- 292 CTG-----ACCTCACCCCGCATATC----- 312
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 TyrTrpAspAspProGluProAspIleGlyAspGlyCysArgThrProGlyGlyArg 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```
Qy 313 ---CTCCATGCTCAGCGATTTTATGTTTCCAGGACCCACCACCAATAATATGGAACAACAT--- 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 104 ArgThrArgLeuTyrAspPheTyrValCysProGly-----HisThr 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 367 -----TCGGAAATCCAGAGATTTCTTTTGTAAACAATGAAGCTGTGTAAACC 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 118 ValProIleGlyCysGlyGlyProGlyGlyTyrCysGlyLysTrpGlyCysGluThr 137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 415 TCTAATGATGATATTGGAATGCGCAACCTCTCAGCAGGATAGGTAAGTTTCTTAT 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 ThrGlyGlnAlaTyrTrpLys---ProSerSerSerTrpAspLeuIleSerLeuLysArg 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 475 GTCAACACC-----TATACCGAGCTCTGGACAAATTTAAATTAACCTG 513
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 157 GlyAsnThrProLysAspGlnGlyProCysTyrAspSerSerValSerSerAsp----- 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 514 ACCTGGATTAGAACTGGAAGCCCC-----AAGTGCCTCTCTTCAGACCTAGATTACCTA 567
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 175 -----IleLysGlyAlaThrProGlyGlyArgCysAsnPro-----Leu 187
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 568 AAAATAAGTTTCACTGAGAAAGGAAACAGAAATAATATCTTAAATGGGTAATGGTATG 627
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 188 ValLeuGluPheThrAspAlaGlyLysLysAlaSer-----TrpAspGlyProLys 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 628 TCTTGGGATGGTATATTATGAGGCTCGGGTAAACAACACCGAGCTCCATTTCTAACTATT 687
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 205 ValTrpGlyLeuArgLeuTyrArgSerThrGlyThrAspProValThrArgPheSerLeu 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 688 CGCCTCAAAATAAACAGCTGGAGCTCCATGCTATAGGACCAAAATACCGTCTTGACG 747
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 225 ThrArgValLeuAsnIleGlyProArgValProIleGlyProAsnProValIleIle 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 748 GGTCAAGACCCCCAACCCCAA----- 768
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 AspGlnLeuProProSerArgProValGlnIleMetLeuProArgProProGlnProPro 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 769 GGACGAGCACCCTCTTAACATAACTTCTGGATCAGACCCCACTGAGCTTAACAGCAGC 828
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 265 ProProGlyAlaAlaSerIleValProGluThrAlaProProSerGlnGlnProGlyThr 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 829 ACTAAATGGGGGCAAACTTTTATGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCC 888
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 285 -----GlyAspArgLeuAsnLeuValAspGlyAlaTyrGlnAlaLeuAsnLeu 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 889 ACGACTCCAGAGCTACTCTTCTTGTGGCTATGCTTAGCTTCGGGCCACCTTACTAT 948
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 302 ThrSerProAspLysThrGlnGluCysTrpLeuValAlaGluProProTyrTyr 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 949 GAAGGAATGGCTAGAGAGGGAATTCATCTGACAAAGAACAATAGAGACCAATGACACA 1008
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 322 GluGlyValAlaValLeuGlyThrTyrSerAsnHisThrSerAlaProThrAsnCysSer 341
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1009 TGGGGATCCCAAAATAAGCTTTACCTTACTGAGGTTTCTGGAAAGGACCTTCGACATAGA 1068
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 342 ValAlaSerGlnHisLysLeuThrLeuSerGluValThrGlyGlnGlyLeuCysIleGly 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1069 AAGGTTCCCATCCCAACACCTTTGTAAACACACTGAAGCCTTTAATCAACACTCT 1128
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 362 ThrValProLysThrHisGlnAlaLeuCysAsnThrThrLeuLysThrAsnLys----- 379
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1129 GAGACTCAATATCTGGTACCTGGTTATGACAGGTGGTGGCATGTAATACTGGATTAAACC 1188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 380 GlySerTyrTyrLeuValAlaProAlaGlyThrThrTrpAlaCysAsnThrGlyLeuThr 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1189 CTTGTGTTTCCACCTGGTTTAAACCAAACTAAAGATTTTTCATTATGGTCCAAAT 1248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 400 ProCysLeuSerAlaThrValLeuAsnArgThrThrAspTyrCysValLeuValGluLeu 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1249 GTTCCCGAGGTATTACTATATCCGAAAGCAATCCTTGTGATGAATATGACTACAGAAAT 1308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 TrpProArgValThrTyrHisProSerTyrValTyrSerGlnPheGlu---LysSer 438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x VCMVCB (1-661)

QY	31	CCGATTCCGGGTGAAAGAGCCGAAAGAGACTGAAATCCCTTAAAGCTTCGGCTCCATCGCG	90
Db	4		22
QY	91	TGGTTCCTTACTCTGTCAAATCACTCTCAGACTAAATGGTATGGCATATGAGAGACACGCTG	150
Db	23		36
QY	151	AACTCCCAATAACCCCTTATCTCTCACCTGGTTAATTACTGACTCCGGCACAGGTATTAAT	210
Db	37		55
QY	211	ATCAACAACACTCAAGGGGAGGCTCCTTTAGGAACCTGGTGGCTGATCTATATAC-----	264
Db	56	ValTTPAlaValThrGlyAsnHisProLeuThrTrpTrpProAspLeuThrProAsp	75
QY	265	GTTTGCTCAGATCAGTT-----ATTCTAGTCGTG	294
Db	76	LeuCysMetLeuAlaLeuHisGlyProThrHisTrpGlyLeuAspAsnHisProTrp	95
QY	295	ACCTCACCCCCAGAT-----	309
Db	96	SerSerProGlyProProCysCysSerGlyAspAlaGlyAlaValSerGlyCysAla	115
QY	309	-----	309
Db	116	ArgAspCysAspGluProLeuThrSerTyrSerProArgCysAsnThrAlaTrpAsnArg	135
QY	310	-----ATCTCCCATGCT-----CACGGATTTTATGTTTGGCCCGACCA	348
Db	136	LeuLysLeuAlaArgValThrHisAlaProLysGluGlyPheTyrIleCysProGlySer	155
QY	349	-----CCAAATAATGAAACAATTCGGGAATCCAGAGATTCTTTTGTAAACATGG	402
Db	156	HisArgProArgTrpAlaArgSerCysGlyLeuAspAlaTyrCysAlaSerTrp	175
QY	403	AACTGTGTAACCTCTAATGATGATATGTGAATGGCAACCTCTCAGCAGATAGGTA	462
Db	176	GlyCysGluThrThrGlyArgAlaAlaTrpAsn---ProThrSerSerTrpAspTyrIle	194
QY	463	AGTTTCTTATGTCAACACTATACCAGCTCTGGACAATTTAATTAACCTGACCTGGATT	522
Db	195	ThrValSer-----AsnAsnLeuThrSerSer-----GlnAlaThrLysAla	208
QY	523	AGAACTGGAGCCCAAGTCTCTCTCTCAGACTAGATTAACCTAAATAATAGTTTCACT	582
Db	209	CysLysAsnAsnGlyTrpCysAsnPro-----LeuValIleArgPheThr	223
QY	583	GAGAAAGGAAACAAGAAATATATCTTAAATGGTAAATGGTATGCTCTGGGGAATGTA	642
Db	224	GlyProGlyLysArgAlaThr-----SerTrpThrThrGlyHisPheTrpGlyLeuArg	241
QY	643	TATTATGGAGCTCGGTAAACAACAGGCTCCATTTCAACTATTCGCCTCAAAATAAAC	702
Db	242	LeuTyr---IleSerGlyHisAspProGlyLeuThrPheGlyIleArgLeuLysValThr	260
QY	703	CAGTGGAGCTCAATGGCTATAGAGCAATAATACGGTCTTCAGCGGGTCAAGACCCCA	762
Db	261	AspLeuGlyProArgValProIleGlyProAsnProValLeuSerAspGlnArgProPro	280
QY	763	ACC-----CAAGGACGAGGCATCTCTAAATAACTTCTCGATCAGAC	807
Db	281	SerArgProValProAlaArgProProProProSerAlaSerProSerThrProThrIle	300
QY	808	CCCACTGAGCTTAACAGACGACTAAATGGGGGCAAAACTTTTATAGCTCATCCAGGA	867
Db	301	ProProGlnGlnGlyThr-----GlyAspArgLeuLeuAsnLeuValGlnGly	316
QY	868	GCTTTTCAAGCTCTTAACCTCCAGACTCCAGAGGCTACCTCTTCTTGTGGCTATGCTTA	927
Db	317	AlaTyrLeuThrLeuAsnMetThrAspProThrArgThrGluGluCysTrpLeuLeuValLeu	336

**Qy** 928 GCTTCGGGCCACCTTACTATGAAGGAATGGCTAGAAGAGGAAATCAATGTGCACAAA 987  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 337 ValSerGluProProTyrrTyrrGluGlyValAlaValLeuArgGluTyrrThrSerHisGlu 356  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 988 GAACATAGAGACCAATGCACATNGGGGATCCCAAANTAGCTTACCCTTACTGAGGTTTCT 1047  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 357 ThrAlaProAlaAsnCysSerSerGlySerGlnHisLysLeuThrLeuSerGluValThr 376  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1048 GGAAAAAGCGCACCTCATAGGAAGGGTCCCCCATCCCACCAACAACCTTTGTAAACCACT 1107  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 377 GlyGlnGlyArgCysLeuGlyThrValProLysThrHisGlnAlaLeuCysAsnArgThr 396  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1108 GAAGCCTTTAATCAAACTCTGAGAGTCAATATCTTGCTTACCTGGTTATTGACAGGTGGTGG 1167  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 397 GluPro-----ThrValSerGlySerAsnTyrrLeuValAlaProGluGlyThrLeuTrp 414  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1168 GCATGTAATACTGGATTAACCCCTTGTTGTTTCCACCTTGGTTTTTAAACCAAACTAAAGAT 1227  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 415 AlaCysSerThrGlyLeuThrProCysLeuSerThrThrValLeuAsnLeuThrThrAsp 434  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1228 TTTTGCATTATGGTCCAAATTTGTTCCCGAGTGTATTACTATCCGGAANAAGCAATCCTT 1287  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 435 TyrCysValLeuValGluLeuTrpProLysValThrTyrrHisSerProAspTyrrValtyr 454  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1288 GATGAATATGACTACAGAAATCATCGCAAAAGAGAGAACCCATATCTCTGACACTTGCT 1347  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 455 ThrGlnPheGlu---ProGlyAlaargPheargargGluProValSerLeuThrLeuAla 473  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1348 GTGATG---CTCGGACTT-----GGAGTGGCAGCAGGTGTAGGAACAGAAACAGCT 1395  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 474 LeuLeuProGluGlyLeuThrMetGlyIleAlaAlaGlyValGlyThrGlyThrThr 493  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1396 GCCCTGTCTACGGGACCA-----CACAGCTAGAAACAGACTTAGTACCTACATCGA 1449  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 494 AlaLeuValAlaThrGlnGlnPheGlnLeuGlnAlaLaMetHisAsn----- 510  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1450 ATTGTAACAGAAGATCTCCAAGCCCTAGAAAAATCTGTCACTAACCTGGAGGAATCCCTA 1509  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 511 -----AspLeuLysGluValGluLysSerIleThrAnLeuGluLysSerLeu 526  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1510 ACTCTCTTATCTGAAGTAGTCTCTACAGAATAGAAGAGGGTTAGATTATTATTCTAAAA 1569  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 527 ThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuLys 546  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1570 GAAGGAGGATTATGTGTAGCCTTCAAGGAGAAAGCTGTTTTTATGTGGATCATTCAGGG 1629  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 547 GluGlyGlyLeuCysAlaAlaLeuLysGluGluCysCysPheTyrrAlaaphHisThrGly 566  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1630 GCCATCAGAGACTCCATCAACCAAGCTTAGACAAAAGTTGGAGAAGCGTCGAAGAGAAAAG 1689  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 567 LeuValIleArgSerMetAlaLysLeuArgGluArgLeuAnGlnArgGlnLysLeuPhe 586  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1690 GAAACTACTCAAGGGTGGTTTGAGGGATGGTTCAACAGGCTCTCTTTGGTTGGCTACCCCTA 1749  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 587 GluSerGlyGlnGlyTyrrPheGluGlyLeuPheAsnArgSerProTyrrPheThrThrLeu 606  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1750 CTTTCTGCTTTAACAGGACCTTAATAGTCTCTCTCTGTTACTCAGAGTTGGGCCATGT 1809  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 607 IleSerThrIleMetGlyProLeuIleValLeuLeuLeuIleLeuLeuPheGlyProCys 626  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1810 ATTTATTAAAGTTAAATTTGCCCTTCATTAGAGAACGAATAAGTCAGCTCCAGATCATGCTA 1869  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 627 IleLeuAsnArgLeuValGlnPheValLysAspArgIleSerValValGlnAlaLeuVal 646  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Qy** 1870 CTTAGACAACAGTAC 1884  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**Db** 647 LeuThrGlnGlnTyrr 651  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**RESULT 15**  
**VGVVFG**  
env polyprotein - feline leukemia virus (strain A/Glasgow-1)  
N:Alternae names: coat polypeptide

N:Contains: coat protein p15E; knob protein gp70  
C:Species: feline leukemia virus  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 12-Apr-1996  
C:Accession: A24300  
R:Stewart, M.A.; Warnock, M.; Wheeler, A.; Wilkie, N.; Mullins, J.I.; Onions, D.E.; Neil  
J. Virol. 58, 825-834, 1986  
A:Title: Nucleotide sequences of a feline leukemia virus subgroup A envelope gene and lo  
A:Reference number: A24300; MUID:86200439; PMID:3009890  
A:Accession: A24300  
A:Molecule type: DNA  
A:Residues: 1-642 <STE>  
A:Cross-references: UNIPARC:UPI0000174A4B  
A:Note: the authors translated the codon GCT for residue 158 as Thr and ACC for residue  
C:Genetics:  
A:Gene: env  
C:Superfamily: type C retrovirus env polyprotein  
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei  
F:1-33/Domain: signal sequence #status predicted <SIG>  
F:34-445/Product: knob protein gp70 #status predicted <GP7>  
F:446-642/Product: coat protein p15E #status predicted <PI1>  
F:35,43,58,91,267,302,374,390,410,578/Binding site: carbohydrate (Asn) (covalent

Alignment Scores:  
Pred. No.: 1,77e-93 Length: 642  
Score: 1272.50 Matches: 282  
Percent Similarity: 57.5% Conservative: 109  
Best Local Similarity: 41.5% Mismatches: 184  
Query Match: 36.7% Indels: 105  
DB: 1 Gaps: 23

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x VCMVPG (1-642)

Qy	4	CATCCCGAGTTAAACCGCGCCACCTCCCGATTCGGGGTGGAAAGCGAAAAGACGTGAAA	63
Db	6	HisPro-----LyseProSerLysAspLysThyLeuSer	16
Qy	64	ATCCCTTAAAGCTTCGCCCTCCATCCGGTGGTTCCTTACTCTGTCATATAACCTCTCAGACT	123
Db	17	TrpAsnLeuAlaPheLeu--ValGlyIleLeuPheThrIleAspIle-----	31
Qy	124	AATGGTATGCGCATAGGAGACGCTGAACTCCCATAAACCCTTAATCTCTACCTGGTTA	183
Db	32	--GlyMetAla-----AsnProSerProHisGlnIleTyrAsnValThrTrpVal	47
Qy	184	ATTACTGACTCGGCACAGGTATTAATATCAACAACACTCAAGGGGAGGCTCTCTTAGGA	243
Db	48	IleThrAsnValGlnThrAsnThrGlnAlaAsnAlaThrSerMetLeuGlyThrLeuThr	67
Qy	244	ACCTGGTGGCTGATCTATACGCTTTGCCCTCAGATCATGTTATT-----CCT	288
Db	68	AspAlaTyrProThrLeuHisValAspLeuValGlyAspThrTrpGluPro	87
Qy	289	AGTCGACTCTACCCCGAGATATCTCCAT-----	318
Db	88	IleValLeuAsnProThrAsnValLysHisGlyAlaArgTyrSerSerLysTyrGly	107
Qy	319	-----GCTCACGGATTTTATGTTTGGCCACGGA	345
Db	108	CysLysThrThrAspArgLysLysGlnGlnThrTyrProPheTyrValCysProGly	127
Qy	346	---CCACCAAAAT-----AATGGAAACATTCGGGAATCCACAGAGATTTCTTTGT	393
Db	128	HisAlaProSerLeuGlyProLysGlyThrHisCysGlyGlyAlaGlnAspGlyPheCys	147
Qy	394	AAACAATGAACTGTGTAACCTCTAATGATGGATATATGGAAATGGCCAACTCTCAGCAG	453
Db	148	AlaAlaTrpGlyCysGluThrThrGlyGluAlaTrpTrpLys---ProThrSer-----	164
Qy	454	GATAGGGTAAGTTTCTTATGTCACACCTATACCAGCTCTCGACAATTAATTAACCTG	513
Db	165	-----SerTrpAspTyrIleThrValLysArgLysSerSerGlnAspAsn-----	179
Qy	514	ACCTGGATTAGAACTGGAGGCCCAAGTGCTCTCTCTCAGACCTAGATTACCTAAAAATA	573

```
Db 180 .....SerCysGluGlyLysCysAsnPro.....LeuValLeu 190
Qy 574 AGTTTCACTGAGAAAGAAAACACAGAAATATCTTAAATCGGTAAATGGTATGCTCTGG 633
Db 191 GlnPheThrGlnLysGlyArgGlnAlaSer.....TrpAspGlyProLysMetTrp 207
Qy 634 GGAATGGTATATATGAGGCTCGGGTAAACACAGGCTCCATTTCAACTATTTCGCTC 693
Db 208 GlyLeuArgLeuTyr--ArgThrGlyTyrAspProIleAlaLeuPheThrValSerArg 226
Qy 694 AAATAAACACAGCTGGAGCTCCATGGCTATAGAGCAATAATAGGCTCTGACGGGTCAA 753
Db 227 GlnValSerThrIleThrProProGlnAlaMetGlyProAsnLeuValLeuProAspGln 246
Qy 754 AGACCCCAACCAAGGACCAAGCACCATCTCTAACATA.....ACT 795
Db 247 LysProProSerArgGlnSerGlnThrGlySerLysValAlaThrGlnArgProGlnThr 266
Qy 796 TCTGGATCAGACCCCACTGAGCTCTAACAGCACGACTAAATGGGGCAAAA-----846
Db 267 AsnGluSerAlaPro--ArgSerValAlaProThrThrMetGlyProLysArgIleGly 285
Qy 847 -----CTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCAGCACT 894
Db 286 ThrGlyAspArgLeuIleAsnLeuValGlnGlyThrTyrLeuAlaLeuAsnAlaThrAsp 305
Qy 895 CCAGAGGCTACCTCTCTTGTGGCTATGCTTAGCTTCGGGCCACCTTACTATGAAGGA 954
Db 306 ProAsnLysThrLysAspCysThrLeuValSerGlyLeuValSerArgProProTyrTyrGluGly 325
Qy 955 ATGGCTAGAGAGGGAATTCATGTGACAAAAGAACATAGAGACCAATGCACATGGGGA 1014
Db 326 IleAlaIleLeuGlyThrTyrSerAsnGlnThrAsnProProProSerCysLeuSerThr 345
Qy 1015 TCCCAAAATAGCTTACCTTACTGAGTTCTGGAAGAGCCACCTGCATAGGAAAGTTT 1074
Db 346 ProGlnHisLysLeuThrIleSerGluValSerGlyGlnGlyMetCysIleGlyThrVal 365
Qy 1075 CCCCATTCCCAACCAACCTTTGTAAACACACCTTAAGCCTTTAATCAACCTCTGAGAGT 1134
Db 366 ProLysThrHisGlnAlaLeuCysAsnLysThrGlnGlnHis-----ThrGlyAla 383
Qy 1135 CAATATCTGGTACCTGGTTATGACAGGTGGTGGCATGTAATCTGGATTAAACCCCTTGT 1194
Db 384 HisTyrLeuAlaAlaProAsnGlyThrTyrTrpAlaCysAsnThrGlyLeuThrProCys 403
Qy 1195 GTTTCACCTTGGTTTTTAACCAAACTAAAGATTTTTCATATGGTCCAAATTTGTTCCC 1254
Db 404 IleSerMetAlaValLeuAsnTrpThrSerAspPheCysValLeuIleGluLeuTrpPro 423
Qy 1255 CGAGTGTTACTATCCCGAAAAGCAATCCTTGATGAATATGACTACAGAAATCAT---1311
Db 424 ArgValThrTyrHisGln-----ProGluTyrValTyrThrHisPheAla 438
Qy 1312 -----CGACAAAAGAGAACCCATATCTCTGACACTTGCTGTGATGCTCGGA---1359
Db 439 LysAlaValArgPheArgArgGluProIleSerLeuThrValAlaLeuMetLeuGlyGly 458
Qy 1360 -----CTTGGAGTGGCAGCAGGTGTAGGAACAGGAACAGCTGCCCTGGTCAACGGA 1410
Db 459 LeuThrValGlyGlyIleAlaAlaGlyValGlyThrGlyThrLysAlaLeu-----475
Qy 1411 CCACAGCAGCTAGAAACAGGA---CTTAGTAACCTACATCGAATTTAAACAGAGATCTC 1467
Db 476 -----LeuGluThrAlaGlnPheArgGlnLeuGlnMetAlaMetHisThrAspIle 492
Qy 1468 CAAGCCCTAGAAAATCTGTCACTAAGTCTGGAGGAATCCCTTAACCTCTTATCTGAAGTA 1527
Db 493 GlnAlaLeuGluSerIleSerAlaLeuGluLysSerLeuThrSerLeuSerGluVal 512
Qy 1528 GTCTCAGAAATAGAGGGTTAGATTATTATTCTTAAAGAGGAGGATTATGTGTA 1587
|||||
```

Search completed: February 14, 2006, 16:07:34  
Job time : 109.883 secs

**This Page Blank (uspto)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 12:52:14 ; Search time 79.6345 Seconds  
(without alignments)  
3391.447 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_5620\_7533  
Perfect score: 3468  
Sequence: 1 ATGCATCCACGTTAAACCG.....CTACACGGGAAGCTGGCGC 1914

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSSWEB.spool/US10723552/runat 14022006 125141 12876/app query.fasta\_1  
-DB=UniProt -QPM=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US10723552 @CGN 1.1 808 @runat 14022006 125141 12876 -NCPU=6 -ICPU=3  
-NO MAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WAEN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80.\*  
1: uniprot.sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3420	98.6	638	2	062707_PIG
2	3407	98.2	638	2	Q90LX4_GAMR
3	3402	98.1	638	2	062705_PIG
4	3402	98.1	638	2	Q90LX3_GAMR
5	3398	98.0	638	2	Q90LX5_GAMR
6	3363	97.0	638	2	Q5J1S9_PIG
7	3293.5	95.0	641	2	Q8Q6V9_GAMR
8	3153.5	90.9	653	2	Q5QG07_GAMR
9	3150	90.8	642	2	Q5J1S8_SUSBA
10	3147.5	90.8	653	2	Q8Q620_GAMR
11	3135.5	90.4	653	2	Q8Q621_GAMR
12	3018.5	87.0	659	2	Q8Q6V8_GAMR
13	2996.5	86.4	659	2	Q8Q6V6_GAMR
14	2995.5	86.4	659	2	Q8Q6V7_GAMR
15	2948.5	85.0	659	2	Q6UQ77_GAMR
16	2943.5	84.9	659	2	Q4VFZ1_GAMR

17	2940.5	84.8	659	2	Q9Q9X3_GAMR
18	2928.5	84.4	659	2	Q6UQ78_GAMR
19	2916.5	84.1	659	2	Q5J1S7_CETA
20	2916.5	84.1	659	2	Q6UQ79_GAMR
21	2866.5	82.7	660	2	Q41172_GAMR
22	2862.5	82.5	660	2	Q8J4V7_GAMR
23	2862.5	82.5	660	2	Q8Q6U6_GAMR
24	2862.5	82.5	2378	2	Q9Q1X4_GAMR
25	2857.5	82.4	660	2	Q6W4U7_PIG
26	2852	82.2	654	2	Q6WGH9_PIG
27	2851	82.2	654	2	Q90RL8_GAMR
28	2845	82.0	654	2	Q6W4U6_PIG
29	2844.5	82.0	660	2	Q6W4V3_PIG
30	2835.5	81.8	645	2	Q6W4U5_PIG
31	2804.5	80.9	661	2	Q9IU66_GAMR
32	2788	80.4	655	2	Q516D0_GAMR
33	2787	80.4	654	2	Q8UM98_GAMR
34	2786.5	80.3	649	2	Q6W4V2_PIG
35	2769.5	79.9	660	2	Q8JEM7_GAMR
36	2765	79.7	654	2	Q6W4V1_PIG
37	2760.5	79.6	661	2	Q9IU67_GAMR
38	2685.5	77.4	612	2	Q6W4U8_PIG
39	2683	77.4	633	2	Q6W4U9_PIG
40	2642.5	76.2	613	2	Q9GLD5_PIG
41	2628	75.8	589	2	Q90LX2_GAMR
42	2541.5	73.3	594	2	Q8WN83_PIG
43	2541.5	73.3	594	2	Q8WN84_PIG
44	2531.5	73.0	594	2	Q8WN77_PIG
45	2523.5	72.8	594	2	Q8WN86_PIG

ALIGNMENTS

RESULT 1  
062707\_PIG  
ID 062707\_PIG PRELIMINARY; PRT; 638 AA.  
AC 062707;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Env protein, [pig].  
OS Sus scrofa, [pig].  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=miniatore swine;  
RX MEDLINE=98216827; PubMed=9557749;  
RA Akiyoshi D.E., Denaro M., Zhu H., Greenstein J.L., Banerjee P., Fishman J.A.  
RT "Identification of a full-length cDNA for an endogenous retrovirus of miniatore swine."  
RL J. Virol. 72:4503-4507(1998).  
DR EMBL; AF038600; AAC16765.1; -, mRNA.  
DR HSP; P03385; IMOF.  
DR SMR; 062707; 488-540.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002050; Env\_polyprotein.  
DR InterPro; IPR008981; FMuLVrecept-bind.  
DR Pfam; PF00429; TLV\_coat; 1.  
SQ SEQUENCE 638 AA; 70997 MW; C8EC46D0D23BE4C8 CRC64;

Alignment Scores:  
Pred. No.: 2,79e-268 Length: 638  
Score: 3420.00 Matches: 637  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.8% Mismatches: 0  
Query Match: 98.6% Indels: 0  
DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x 062707\_PIG (1-638)

```
QY 1 ATGCATCCACGTTAAACCGGGCCACCTCCCGATTCCGGGTGGAAACCCGAAAGACTG 60
Db 1 MethHisProThrLeuAsnArgHisLeuProIleArgGlyGlyLeuProLysArgLeu 20
QY 61 AAAATCCCTTAAGCTTCGCTCCCATCGGTGGTTCCTTACTCTCTCAATAACCTCTCAG 120
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40
QY 121 ACTAATGTTATGCGATAGGACAGCCTGAACCTCCCATAAACCTTATCTCTCACCTGG 180
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTACTGACTCCGGCAGACAGTAAATATCAACAACACTCAAGGGAGGCTCTTTA 240
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80
QY 241 GGAACTCGTGGCTGATCTATACGTTTGCCTCAGATCAGTTATTCCTAGTCTGACCTCA 300
Db 81 GlyThrTrpTrpProAspLeuTrpValCysLeuArgSerValIleProSerLeuThrSer 100
QY 301 CCCCAGATATCCTCCATCTCACGGATTTTATGTTTCCCGGAGCACCAACAAATAATGGA 360
Db 101 ProProAspIleLeuHisAlaHisGlyPheTrpValCysProGlyProProAsnAsnGly 120
QY 361 AAACATTCGGGAATCCAGAGATTTCTTTGTGTAACAATGGAACGTGTAACTCTTAAT 420
Db 121 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 140
QY 421 GATGATATTTGAAATGGCCAACTCTCAGCAGGATAGGTAAGTTTCTTATGTCTAAC 480
Db 141 AspGlyTrpTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTrpValAsn 160
QY 481 ACCTATACCACTCTGGCAATTTAATTAACCTGACCTGGATTTAGAACTGGAGCCCAAG 540
Db 161 ThrTrpThrSerSerGlyGlnPheAsnTrpLeuThrTrpIleArgThrGlySerProLys 180
QY 541 TGCTCTCTTCAGACTAGATTACCTAAATAAGTTTCTACCTGAGAAAGGAAACAAGAA 600
Db 181 CysSerProSerAspLeuAspTrpLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200
QY 601 AATATCTTAAATGGTAAATGGTATGCTTGGGAATGGTATATTATGGAGGCTCGGGT 660
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTrpTrpGlySerGly 220
QY 661 AAACAACCAAGCTCCATCTTAATCTTCGCTCCCAATAAACCAGCTGGAGCTCCAATG 720
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProMet 240
QY 721 GCTATAGACCAATACGGCTTTGACGGGTCAAGACCCCAACCCCAAGGACCGAGGACCA 780
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260
QY 781 TCCTCTAACATAACTTCGGATCAGACCCCACTAGTCTTAACAGCACCACTAAATATGGG 840
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerSerSerThrThrLysMetGly 280
QY 841 GCAAACTTTTATAGCTCATCAGGGAGCTTTTCAAGCTCTTAACTCCACCACTCCAGAG 900
Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300
QY 901 GCTACTCTTCTTGTGTAGTCTAGCTTCGGGCCCACTTACTTACTAAGAGGAATGGCT 960
Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTrpTrpGlyMetAla 320
QY 961 AGAAGAGGAATTCATGTGCAAAAGAACATAGACCAATGACATGCGGATGGGATCCCAA 1020
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpLysSerGln 340
QY 1021 AATAAGCTTACCTTACTGAGTTTCTCGAAAGGACCTGCTGATAGGAAGGTTCCCCCA 1080
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360
```

```
QY 1081 TCCCAACCAACACCTTTGTAAACCACTGAAGCCCTTAAATCAAACTCTGAGAGTCAATAT 1140
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTrp 380
QY 1141 CTGTACTCCTGTTATGACAGGTGGTCATGTAAATACTGATTAAACCCCTTGTGTTTCC 1200
Db 381 LeuValProGlyTrpAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400
QY 1201 ACCTTGGTTTTAAACCAAACTAAAGATTTTGCATTATGTCGTCCAAATTTGTTCCCGAGTG 1260
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420
QY 1261 TATTACTATCCGAAAGCAATCCTTCATGAATATGACTACAGAAATCATCGCAAAAG 1320
Db 421 TyrTrpTrpProGluLysAlaIleLeuAspGluTrpAspTrpArgAsnHisArgGlnLys 440
QY 1321 AGAACAACCATATCTCTGACACTTGTGTGATGTCGACTTGGAGTGGCAGCAGGTGTA 1380
Db 441 ArgGluProLysSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460
QY 1381 GGAAACAGAACAGCTGCCCTGGTCAACGGGACCAACAGCAGCTAGAAACAGGACTTAGTAAC 1440
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
QY 1441 CTACATCGAATTGTAAACAGAGATCTCCAAGCCCTAGAAAATCTGTCAGTAACTCGGAG 1500
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500
QY 1501 GAATCCCTTAACTCTTATCTGAAGTAGTCTCAGCAATAGAAAGGAGGTTAGATTATTA 1560
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeu 520
QY 1561 TTTCTAAAGAGAGAGATTATGTAGCTTGAAGGAGGAATGCTGTTTATTTATGTGGAT 1620
Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTrpValAsp 540
QY 1621 CATTCAGGGCCATCAGAGACTCCATGAACAGCTTAGAGAAAGCTTCGAGAAGCGTCGA 1680
Db 541 HisSerGlyAlaIleLeuArgAspSerMetAsnLysLeuArgGluArgGluLysArgArg 560
QY 1681 AGGAAAAGGAAACTACTCAAGGGTGGTTTCAGGGATGGTTCAACAGACTCTCTTTGGTTG 1740
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeu 580
QY 1741 GCTACCTTACTTCTGCTTTAAACAGGACCCCTTAATAGTCCCTCTCTCTTACTCACAGTT 1800
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600
QY 1801 GGGCATGTATTATTAAACAGTTAATTCCTTCAATAGAGAACGAATAAGTGCAGTCCAG 1860
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgGileSerAlaValGln 620
QY 1861 ATCATGTTACTTAGACAACAGTACCAAGCCCGCTCTAGCAGGAGAGCTGGCGCG 1914
Db 621 IleMetValLeuArgGlnGlnTrpGlnSerProSerSerArgGluAlaGlyArg 638
```

## RESULT 2

```
Q90LX4_9GAMR PRELIMINARY; PRT; 638 AA.
AC Q90LX4_
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope protein.
OS Porcine endogenous type C retrovirus.
OC Viruses; Retroid viruses; Retroviridae; Gammatetrovirus.
OX NCBI_TaxID=70540;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Blusch J.H., Seelmeier S., von der Helm K.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF402661; AAK94076.1; -; Genomic_DNA.
```

QY	781	TCCTCTAACATAACTTCTGGATCAGACCCCACTGAGTCTAACACGACGACTAAATAATGGG	840
DB	261	SerSerAsnIleThrSerGlySerAspProThrGluSerAenSerThrThrLysMetGly	280
QY	841	GCRAAACTTTTATAGCCTCATCCAGGGAGCTTTCAAGCTCTTAAGTCTTAACTCCACGACTCCAG	900
DB	281	AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAenSerThrThrProGlu	300
QY	901	GCTACCTCTCTTCTTGGCTATGCTAGCTTCGGGCCACCTTACTATGAAGGAATGGCT	960
DB	301	AlaThrSerSerCysTyrLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAla	320
QY	961	AGAAGAGGGAATTCATATGTGCAAAAAGAACATAGAGACCAATGCATCGGGGATCCCAA	1020
DB	321	ArgArgGlyLysPheAenValThrLysGluHisArgAspGlnCysThrTyrGlySerGln	340
QY	1021	AATAAGCTTACCTTACTGAGGTTTCTGGAAGAAGGCACCTGCATAGAAAGGTTCCCCCA	1080
DB	341	AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro	360
QY	1081	TCCCAACCAACACCTTTGTAACCAACACTGAAGCCCTTTAATCAAACTCTGAGAGTCAATAT	1140
DB	361	SerHisGlnHisLeuCysAenHisThrGluAlaPheAenGlnThrSerGluSerGlnTyr	380
QY	1141	CTGCTACCTCGTTATGACAGGTGGTGGCATGTAAATCTGGATTAACCCCTTGTTTCC	1200
DB	381	LeuValProGlyTyrAspAsgTyrTpAlaCysAenThrGlyLeuThrProCysValSer	400
QY	1201	ACCTTGGTTTTTAACCAACTAAAGATTTTTTGCATTTATGCTCMAATGTTTCCCGAGTG	1260
DB	401	ThrLeuValPheAenGlnThrLysAspPheCysIleMetValGlnIleValProArgVal	420
QY	1261	TATTACTATCCGAAAAGCAATCCTTGATGAATATGACTACAGAATACTCAGCAAAAG	1320
DB	421	TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspIlyrArgAenHisArgGlnLys	440
QY	1321	AGAAACCCATATCTCTGACACTTGTGTGATGCTCGGACTTGGAGTGGCAGCAGGTGA	1380
DB	441	ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal	460
QY	1381	GGACAGGACAGCTGCCCTGGTTCACGGGACCAAGCAGCTAGAAACAGACACTTAGTAAC	1440
DB	461	GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn	480
QY	1441	CTACATCGAATGTAAACAGAAAGATCTCAAGCCCTAGAAAAATCTGTCACTACCTCGAG	1500
DB	481	LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAenLeuGlu	500
QY	1501	GAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGAAATAGAGAGGTTAGATTATTA	1560
DB	501	GluSerLeuThrSerLeuSerGluValValLeuGlnAenArgGlyLeuAspLeuLeu	520
QY	1561	TTTCTAAAGAGGAGGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTGGAT	1620
DB	521	PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAsp	540
QY	1621	CATTTCAGGGCCATCAGAGACTCCATGAACAAGCTTTAGAGAAAGTTGGAGAGCGTCGA	1680
DB	541	HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArg	560
QY	1681	AGGGAAAGGAAACTACTCAAGGCTGGTTTGAAGGATGGTTTCAACAGGCTCTCTTGGTTG	1740
DB	561	ArgGluLysGluThrThrGlnGlyTyrPheGluGlyTyrPheAenArgSerProTyrLeu	580
QY	1741	GCTACCTCTTCTGTTTAAACAGACCTTAAATAGTCTCTCTCTGTTTACTCAGATT	1800
DB	581	AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal	600
QY	1801	GGGCCATGATTATTAAACAGTTAAATGGCTTCATTAGAGAACGAATAAGTCAGTCCAG	1860
DB	601	GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln	620
QY	1861	ATCATGGTACTTAGACAAACAGTACCAAGACCCCTCTAGCAGGGAAGCTGGCCGC	1914

Pred. No.:	3.16e-267	Length:	638
Score:	3407.00	Matches:	635
Percent Similarity:	99.7%	Conservative:	1
Best Local Similarity:	99.5%	Mismatches:	2
Query Match:	98.2%	Indels:	0
DB:	2	Gaps:	0

Qy	1	ATGCATCCACGCTTAACCGCGCCACCTCCCGATTCCGGGTGAAACCGCGAAAGACTG	60
Db	1	MetHisProThrLeuSerArgHisLeuProIleArgGlyGlyLysProLysArgLeu	20
Qy	61	AAAATCCCCTTAAGCTTCGCCTCCATCCGGTGGTTCCTTACTCTGTCAATAACTCTCAG	120
Db	21	LysIleProLeuSerPheAlaSerIlealaTrpPheLeuThrLeuSerIleThrSerGln	40
Qy	121	ACTAATGGTATCGCATAGGACAGCCTGAACCTCCCAATAAACCCCTTATCTCTCACCTGG	180
Db	41	ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp	60
Qy	181	TTAATTACTGATCCCGCACAGGTATTAAATATCAACAACACTCAAGGGAGGCTCCTTTA	240
Db	61	LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu	80
Qy	241	GGAACTGTGGCCGTGATCTATACGTTTGCCTTCAGATCAGTTATTCTAGTCTGACCTCA	300
Db	81	GlyThrTrpTrpProAspLeuTrpValCysLeuArgSerValIleProSerLeuThrSer	100
Qy	301	CCCCAGATATCCTCATGCTCACGGATTTTATGTTGCCCGCAGGACACCAAAATAATGGA	360
Db	101	ProProAspIleLeuHisAlaHisGlyPheTrpValCysProGlyProProAsnAsnGly	120
Qy	361	AAACATTGGGAAATCCCGAGAGATTCTTTTGTAAACAATGGAACTGTGAACCTCTAAT	420
Db	121	LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn	140
Qy	421	GATGGATATTGGAATGGCCACCTCTCAGCAGCATAGGCTAAGTTTTCTTATGTCAAC	480
Db	141	AspGlyTyTrpLysTrpProThrSerGlnAspArgValSerPheSerTrpValAsn	160
Qy	481	ACCTATACCAGCTCTGGCAATTTAATTACCTGACCTGGATTAGAACTGGAAGCCCAAG	540
Db	161	ThrTyThrSerSerGlyGlnPheAsnTyTrpLeuThrTrpIleArgThrGlySerProLys	180
Qy	541	TGCTCTCTTCAGACTGATTAACCTAAAAATAAGTTTCTACTGAGAAAGGAAAAACAAGAA	600
Db	181	CysSerProSerAspLeuAspTrpLeuLysIleSerPheThrGluLysGlyLysGlnGlu	200
Qy	601	AAATACCTAAATGGTAAATGGTATGCTCTGGGGAAATGCTATATATATGAGAGCTCGGT	660
Db	201	AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyTrpGlyGlySerGly	220
Qy	661	AAACAACACGAGCTCCATTTCAACTATTCCGCTCAAAATAAACACAGCTGGAGCCTCCAATG	720
Db	221	LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet	240
Qy	721	GCTATAGGACAAATACGGCTTTGACGGGTCAAGACCCCCAACCCAGGACGAGGACCA	780
Db	241	AlaIleGlyProAsnThrValLeuThrGlyGlnArgProThrGlnGlyProGlyPro	260

```
Db 621 ileMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 638
|||||
RESULT 3
O62705_PIG PRELIMINARY; PRT; 638 AA.
AC O62705
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Env protein.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
NCBI_TaxID=9823;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98216827; PubMed=9557749;
RA Akiyoshi D.E.; Denaro M., Zhu H., Greenstein J.L., Banerjee P.,
RA Fishman J.A.;
RT "Identification of a full-length cDNA for an endogenous retrovirus of
RL J. Virol. 72:4503-4507(1998).
DR EMBL; AF038599; AAC16762.1; -; mRNA.
DR HSP; P03385; IMOF.
DR SMR; O62705; 488-540.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPRO02050; Env polyprotein.
DR InterPro; IPRO08981; FMuLVrecept-bind.
DR Pfam; PF00429; TLV_coat; 1.
SQ SEQUENCE 638 AA; 70980 MW; CFC394814FD23FEA CRC64;

Alignment Scores:
Pred. No.: 8,04e-267 Length: 638
Score: 3402.00 Matches: 634
Percent Similarity: 99.7% Conservative: 2
Best Local Similarity: 99.4% Mismatches: 2
Query Match: 98.1% Indels: 0
DB: 2 Gaps: 0

US-10-723-552-3_COPY_5620_7533 (1-1914) x O62705_PIG (1-638)
QY 1 ATGCATCCACGTTAAACGGGGCCACCTCCGATTCGGGTGGNAACCGAAGACTG 60
Db 1 MetHisProThrLeuSerArgHisLeuProIleArgGlyGlyLeuProLysArgLeu 20
QY 61 AAAATCCCTTAAGCTTCGCTCCATCGCTGGTTCTTACTCTGTCATAACCTCTCAG 120
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40
QY 121 ACTAATGGTATGCGCATAGGACAGCCTGAACCTCCCATAAACCTTATCTCTCACCTGG 180
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTAAGTACGTCGGCAGCAGGATTAAATATCAACAACACTCAAGGGAGGCTCTTTA 240
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTTCAGATCAGTATTTCCTAGTCTGACCTCA 300
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100
QY 301 CCCCAGATATCCTCCATGCTCAGCGATTTTATGTTTGGCCAGGACCAACCAATAATGGA 360
Db 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120
QY 361 AAACATTGGGAAATCCAGAGATTTCTTTGTAAACAAATGAACTGTAACTCTAAT 420
Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140
QY 421 GATGGATATTGGAATGGCCAACTCTCAGCAGGATAGGTAAGTTCCTTCTTATGTCAC 480
```

```
Db 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160
QY 481 ACCTATACAGCTCTGGACAAATTTAACTACCTGACCTGGATTAGAACTGGAGCCCAAG 540
Db 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180
QY 541 TGCCTCTCTTCAGACCTAGATTACTAAATAATAGTTTCTACTGAGAAAGGAAAAACAAG 600
Db 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200
QY 601 AATATCTCTAAATGGTAAATGGTATGTTCTTGGGGAATGGTATATTATGGAGGCTCGGGT 660
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGly 220
QY 661 AAACAACAGGCTCCATCTTAACCTATTTCGCCTCAAAATAAACACAGCTGGAGCCTCCAATG 720
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240
QY 721 GCTATAGGACCAATACGGTCTTCACGGGTCAAGACCCCAACCCCAAGGACCAAGACCA 780
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260
QY 781 TCCTCTAACATAACTTCTGGATCAGACCCCACTGAGTCTAAACAGCACGACTAAATAATGGG 840
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerSerSerThrThrLysMetGly 280
QY 841 GCAAACTTTTAAAGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCACGACTCCAGAG 900
Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300
QY 901 GCTACCTCTCTTCTGGTATGCTAGCTTCGGGGCCACCTTACTATGAAGGAATGGCT 960
Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAla 320
QY 961 AGAAGAGGGAATTCATGTGACAAAGAAACATAGACACCAATGACATGGGGATCCCAA 1020
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340
QY 1021 AATAAGCTTACCTTACTAGGTTTCTGGAAAGGACCTGCATAGGAAAGGTTCCCCCA 1080
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360
QY 1081 TCCCAACACACCTTTGTAAACCACTGAACCTTTTAATCAAACTCTGAGAGTCAATAT 1140
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380
QY 1141 CTGGTACCTGTTATGACAGGTGGTGGGCATGTAAATACTGGATTAAACCCCTTGTGTTCC 1200
Db 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400
QY 1201 ACCTTGGTTTTTAACCAAACTAAAGATTTTTCGATTTATGGTCCAAATTTGTTCCCGAGTG 1260
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420
QY 1261 TATTACTATCCGAAAAGCAATCCTTGATCAATATGACTACAGAAATCATCGACAAAG 1320
Db 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440
QY 1321 AGAGAACCATATCTCTGACACTTTCGCTGTGATGCTCGGACTTGGAGTGGCAGCAGTGTA 1380
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460
QY 1381 GGAAACAGAAACAGCTGCCCTGGTCCAGGGACCAAGAGCTAGAGAAACAGGACTTAGTAAC 1440
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
QY 1441 CTACATCGAATTGTAAACAGAGAGACTCCCAAGCCCTAGAAAAATCTGTCAAGTAACCTGAG 1500
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500
QY 1501 GAATCCCTTAACCTCTTATCTCTGAAGTAGTCTCTACAGATAGAGAGGTTAGATTATTA 1560
|||||
```

Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
 QY 1561 TTTCTAAAGAGGAGGATTATGTAGCTTGAAGGAGGAATGCTGTTTTATGTGGAT 1620  
 Db 521 PheLeuLysGluGlyLeuCysValAlaLeuLysGluCysPheTyrValAsp 540  
 QY 1621 CATTGAGGGCCATCAGAGCTCCATGAACAGCTTAGAGAAAGTTGAGAGCGTCA 1680  
 Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560  
 QY 1681 AGGAAAAGAGAACTACTCAAGGGTGGTTGAGGATGTTCAACAGGTCTCTTTGGTTG 1740  
 Db 561 ArgGluLysGluThrThrGlnGlyTyrPheGluGlyTyrPheAsnArgSerProTyrLeu 580  
 QY 1741 GCTACCTTACTTCTGCTTTTAAAGACCCCTTAATAGTCTCTCTCTCTCTCTCTCT 1800  
 Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
 QY 1801 GGGCCATCTATTATTAAACAGTTAATGCTTCAATTAGAGAACGAATAAGTGCAGTCCAG 1860  
 Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
 QY 1861 ATCATGTACTTAGACACAGTACCMAAGCCCGTCTAGCAGGGAAGCTGGCCGC 1914  
 Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 638

## RESULT 4

Q90LX3\_9GAMR  
 ID Q90LX3\_9GAMR PRELIMINARY; PRT; 638 AA.  
 AC Q90LX3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Envelope protein.  
 OS Porcine endogenous type C retrovirus.  
 OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.  
 OX NCBI\_TaxID=70540;  
 RN (1)  
 RP NUCLEOTIDE SEQUENCE.  
 RA Blusch J.H., Seelmeier S., von der Helm K.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF402662; AAK94077.1; -, Genomic\_DNA.  
 DR HSP; P03385; IMOF.  
 DR SMR; Q90LX3; 488-540.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR002050; Env\_polyprotein.  
 DR InterPro; IPR008981; FmULVrecept-bind.  
 DR Pfam; PF00429; TLV\_coat; 1.  
 KW Envelope protein.  
 SQ SEQUENCE 638 AA; 70980 MW; CFC394814FD23FEA CRC64;

## Alignment Scores:

Pred. No.: 8,04e-267 Length: 638  
 Score: 3402.00 Matches: 634  
 Percent Similarity: 99.7% Conservative: 2  
 Best Local Similarity: 99.4% Mismatches: 2  
 Query Match: 98.1% Indels: 0  
 DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x Q90LX3\_9GAMR (1-638)

QY 1 ATGCATCCACAGTTTAAACCGCGCCACCTCCCGATTCGGGGTGGAAAGCCGAAAGACTG 60  
 Db 1 MeHisProThrLeuSerArgGHisLeuProIleArgGlyLysProLysArgLeu 20  
 QY 61 AAAATCCCTTAAGCTTCGCTCCATCGCGTGGTTCTTACTCTGTCAATAAACCCTTCAG 120  
 Db 21 LysIleProLeuSerPheAlaSerIleAlaTyrPheLeuThrLeuSerIleThrSerGln 40  
 QY 121 ACTAATGGTATGGCATAGAGACAGCTCGAATCCCATAAACCCTTATCTCTCACTGG 180

Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
 QY 181 TTAATTACTGCTCCGGCACAGGTATTAAATCAACACACTCAAGGGGAGGCTCCTTTA 240  
 Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnThrGlnGlyGluAlaProLeu 80  
 QY 241 GGAACTGTGGCTGTATCTATACGTTTGCCTCAGATCAGTTATTCCTAGTCTGACCTCA 300  
 Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
 QY 301 CCCCAGATATCCTCCATGCTCAGGATTTTATGTTTCCCGAGACACCAAAATAATGGA 360  
 Db 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120  
 QY 361 AAAATTTGCGGAATCCAGAGATTTCTTTTGTAAACAATGGAACCTGTGTAACCTCTAAT 420  
 Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140  
 QY 421 GATGGATATTGGAAATGGCCAACTCTCAGCAGATAGGGTAAGTTTCTTATGTCAAC 480  
 Db 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
 QY 481 ACCTATACCACTCTGACAAATTAATTAATCTGACCTGGATTAGAACTGGAGCCCCAAG 540  
 Db 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180  
 QY 541 TGCTCTCTCTCAGACCTAGATTACTAAAATAAGTTTCACTGAGAAGGAAACAAAGAA 600  
 Db 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
 QY 601 AATATCTTAAATGGGTAAATGGTATGCTTTGGGAAATGGTATATATGAGGAGCTCGGGT 660  
 Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGly 220  
 QY 661 AAACAACCAAGCTCTCAATCTTAATTCGCTGCTCAAAATAAACAGCTGGAGCTCCCAATG 720  
 Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
 QY 721 GCTATAGGCAAAATACGGTCTTGACGGGTCAAGACCCCAACCCCAAGGACAGGACCA 780  
 Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
 QY 781 TCCTCTAACATAACTTCTGGATCAGACCCCACTGAGTCTAACACAGCAGCTATAAATGGG 840  
 Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerSerThrThrLysMetGly 280  
 QY 841 GCAAAACCTTTTAGCCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCAGACTCCAGAG 900  
 Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
 QY 901 GCTACCTCTTCTTTGGCTATGCTTACGCTTCGGGCCACCTTACTATGAGGAATGCT 960  
 Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAla 320  
 QY 961 AGAAGAGGGAATTCATATGTGACAAAAGAACATAGAGACCAATGCATATGGGATCCCAA 1020  
 Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
 QY 1021 AATAAGCTTACCTTACTGAGGTTCTGGAAAAGGACCTGCATAGAAAAGTTTCCCCCA 1080  
 Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
 QY 1081 TCCCACCAACACCTTTGTAACCACTGAGCCCTTTAATCAAACTCTGAGAGCTCAATAT 1140  
 Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
 QY 1141 CTGGTACCTGGTTATCAGAGGTGGTGGCATGTAATCTGGAATTAACCCCTTGTGTTTC 1200  
 Db 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
 QY 1201 ACCTTGGTTTTTAACCAAACTAAAGATTTTGCATTATGGTCCAAATGTTTCCCCGAGTG 1260  
 Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420

```
QY 1261 TATTACTATCCGAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAAG 1320
Db 421 TyrtYrtyrProGluLysAlaIleuAspGluYrAspTyrArgAsnHisArgGlnLys 440
QY 1321 AGAGAACCATATCTCTGACACTTGCTGCTGTGATGCTCGGACTTGGAGTGGCAGCAGGTGTA 1380
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460
QY 1381 GGAACAGAAACAGCTGCCCTGTCACGGGACACACAGCAGCTAGAAACAGGACTTAGTAAC 1440
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
QY 1441 CTACATCGAATTGTAAACAGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCTGGAG 1500
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500
QY 1501 GAAATCCCTAAACCTCTTATCTCAAGTAGTCTCTACAGAAATAGAGAGGTTAGATTTATTA 1560
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgAGGlyLeuAspLeuLeu 520
QY 1561 TTTCTAAAGAGGAGGATTAATGTAGCTTGGAAGGAGGAATGCTGTTTTTATGTGGAT 1620
Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAsp 540
QY 1621 CATTGAGGGCCATCAGAGACTCCATGAACAGCTTAGAGAAAGGTTGGAGAGCGTGA 1680
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgarg 560
QY 1681 AGGGAAAAAGGAAACTACTCAAGGGTGGTTTCAGGGATGGTTCAACAGAGTCTCTTTGGTTG 1740
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580
QY 1741 GCTACCCCTACTTTCGCTTTAACAGGACCCCTTAATAGTCCTCCTCCTGTTACTCACAGTT 1800
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600
QY 1801 GGGCCATGATTTATTAAACAGTTAATTCGCTTCAATAGAGAACGTAAGTGCAGTCCAG 1860
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgGluLeuAlaValGln 620
QY 1861 ATCATGTGACTTAGACAACAGTACCAGGCCGCTAGCAGGAGAGCTGGCCGC 1914
Db 621 IleMetValLeuArgGlnGlnTyrglnSerProSerSerArgGluAlaGlyArg 638
```

RESULT 5

```
Q90LX5_9GAMR
ID Q90LX5_9GAMR PRELIMINARY; PRT; 638 AA.
AC Q90LX5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope protein.
OS Porcine endogenous type C retrovirus.
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=70540;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Blusch J.H., Seelmeier S., von der Helm K.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF402660; AAK94075.1; -; Genomic_DNA.
DR HSP; P03385; 1MOP.
DR SMR; Q90LX5; 488-540.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env polypeptide.
DR InterPro; IPR008981; FMuLVrecept-bind.
DR Pfam; PF00429; Tlv_coat; 1.
KW Envelope protein.
SQ SEQUENCE 638 AA; 71053 MW; 15B3A4BCDA40B3C7 CRC64;
```

Alignment Scores:

```
Pred. No.: 1.7e-266 Length: 638
Score: 3398.00 Matches: 633
Percent Similarity: 99.4% Conservative: 1
Best Local Similarity: 99.2% Mismatches: 4
Query Match: 98.0% Indels: 0
DB: 2 Gaps: 0
US-10-723-552-3_copy_5620_7533 (1-1914) x Q90LX5_9GAMR (1-638)
QY 1 ATGCATCCCAAGTTAAACCGCGCCACCTCCGATTCGGGGTGAAGAGCGAAAGACTG 60
Db 1 MethisProthrLeuSerArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20
QY 61 AAAATCCCTTAACTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCAATAACCTCTCAG 120
Db 21 LysileProleuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40
QY 121 ACTAATGTGTATGCGCATAGGAGACAGCTGAACCTCCATAAACCTTTATCTCTCACCTGG 180
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTAAGTCTCCGCGACAGGTATTAAATATCAACACTCAAGGGAGGCTCCTTTA 240
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80
QY 241 GGAACCTGCTGGCTGATCTATACGTTTGCTCAGATCAGTATTCTTAGTCTCACCTCA 300
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100
QY 301 CCCCAGATATCTCCATGCTCACGATTTTATGTTTCCCGCAGGACCAACCAATAATGGA 360
Db 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120
QY 361 AAACATTCGCGAAATCCAGAGATTTCTTTTCTAAACAATGGAACCTGTAACTCTTAAT 420
Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140
QY 421 GATCGATATTGGAATGGCAACCTCTCAGCAGATAGGGTAAGTCTTTCTTATGTCAAC 480
Db 141 AspGlyTyrrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrrValAsn 160
QY 481 ACCTATACCAAGCTCTGGACAAATTTAATTAACCTGACCTGGATTTAGAACCTGCAAG 540
Db 161 ThrTyrrThrSerSerGlyGlnPheAsnTyrrLeuThrTrpIleArgThrGlySerProLys 180
QY 541 TGCTCTCTTCAGACCTAGATTACTTAAATAATAGTTTCACTGAGAAAGGAAAACAGAA 600
Db 181 CysSerProSerAspLeuAspTyrrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200
QY 601 AATATCTTAAATGGGTAAATGGTATGTTCTTGGGGAATGGTATATTATGAGGCTCGGGT 660
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrrGlyLysGlySerGly 220
QY 661 AAACAACCAAGCTCCATTTCAACTATTTCGCTCAAAATAAACACCAAGCTGGAGCCCTCAATG 720
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240
QY 721 GCTATAGACCAATAATCGCTTTCAGGGTCAAGACCCCAACCCCAAGGACCAAGGACCA 780
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260
QY 781 TCCTCTAACATAACTTCTGGATCAGCCCACTCAGTCTAACAGCACCACCTAAATATGGG 840
Db 261 SerCysAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280
QY 841 GCAAAACCTTTTATGCTCATCATCAGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAG 900
Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300
QY 901 GCTACCTCTTCTTGTGGCTATGCTTAGCTTCGGGGCCACCTTACTATGAGAGGATGGCT 960
Db 301 ThrThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrrTyrrGlyMetAla 320
```

```
QY 961 AGAAGCGGAATTCATGTGACAAAGAACATAGAGACCAATGACATGGGATCCAA 1020
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340
QY 1021 AATAAGCTTACCTTACTGAGGTTTCTGAAAGGCCCTGCATAGGAAAGTTCCTCCCA 1080
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360
QY 1081 TCCCAACCAACCTTTGTAACCACTGAAGCCTTTAATCAAACTCTGAGAGTCAATAT 1140
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTrp 380
QY 1141 CTGGTACCTGTTATGACAGGTGGTGGCATGTAATCTGATTAACCCCTTGTGTTCC 1200
Db 381 LeuValProGlyTyArgTrpAlaCysAsnThrGlyLeuThrProCysValSer 400
QY 1201 ACCTTGTTTAAACCAACTAAAGATTTTGCATTATGTCCTCAATGTCCTCCCGAGTG 1260
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420
QY 1261 TATTACTATCCGAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAG 1320
Db 421 TyrTyTrpProGluLysAlaIleLeuAspGluTyArgAsnHisArgGlnLys 440
QY 1321 AGAGAACCATATCTGACACTGCTGTGATGCTCGGACTTGAGTGGCAGCGGTGTA 1380
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460
QY 1381 GGACAGGAACAGCTGCCCTGTCACGGGACACAGCAGCTAGAAACAGGACTTAGTAAC 1440
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
QY 1441 CTACATCGAATTTGAACAGAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACTGGAG 1500
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500
QY 1501 GAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGATAGAGAGGGTTAGATTATA 1560
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520
QY 1561 TTTCTAAAGAGGAGGATTATGTAGCCTTGAAGGAGGAATCTGTTTTTATGTGAT 1620
Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysPheTyValAsp 540
QY 1621 CATTCAGGGCCATCAGAGACTCCATGAACAAAGCTTAGAGAAAGGTTGGAGAACGTCGA 1680
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuLysArgArg 560
QY 1681 AGGAAAGGAACTACTCAAGGTGGTTGAGGGATGGTTCAACAGGTCTCTTTGGTTG 1740
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580
QY 1741 GCTACCTACTCTTCTGTTTAAAGGACCTTATAGTCTCTCTCTCTCTCTCTCTCTCT 1800
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600
QY 1801 GGGCCATGTATTATTAAAGTTAATTGCCCTTCAATTAGAACCAATAGTGACAGTCCAG 1860
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620
QY 1861 ATCATGTACTTAGACACAGTACCAAGCCCGTCTAGCAGGGAAGCTGGCCGC 1914
Db 621 IleMetValLeuArgGlnGlnTyArgGlnSerProSerSerArgGluAlaGlyArg 638
```

RESULT 6

```
Q5J1S9_PIG
ID Q5J1S9_PIG PRELIMINARY; PRT; 638 AA.
AC Q5J1S9;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Envelope glycoprotein.
GN Name=env;
```

```
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OX NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15596862; DOI=10.1128/JVI.79.1.649-654.2005;
RA Niebert M., Tonjes R.R.;
RT "Evolutionary spread and recombination of porcine endogenous
RT retroviruses in the suiformes.";
RL J. Virol. 79:649-654 (2005).
DR EMBL; AY534304; AAT0256.1; -; mRNA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; tlv_coat; 1.
KW Envelope protein.
SQ SEQUENCE 638 AA; 71021 MW; 9428D75C6F21B3B8 CRC64;
```

Alignment Scores:

Pred. No.:	1.17e-263	Length:	638
Score:	3363.00	Matches:	627
Percent Similarity:	98.7%	Conservative:	3
Best Local Similarity:	98.3%	Mismatches:	8
Query Match:	97.0%	Indels:	0
DB:	2	Gaps:	0

US-10-723-552-3\_copy\_5620\_7533 (1-1914) x Q5J1S9\_PIG (1-638)

QY 1 ATGCATCCACGTTAAACCCGCGCCACCTCCCGATTCCGGGTGGAAAGCCGAAAGACTG 60

Db 1 MethiSprothrLeuAsnArgArgHisLeuProIleArgGlyLysProLysArgLeu 20

QY 61 AAAATCCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTGTCTCAATAACCTCTCAG 120

Db 21 LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40

QY 121 ACTAATCGTATCGCATAGGAGACAGCTGAACTCCCAATCAACCTTATCTCTCACCTGG 180

Db 41 ThrAsnGlyMetArgIleGlyAsnSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60

QY 181 TTAATTACTGACTCCGCGCACAGGTATTAAATATCAACAACACTCAAGGGAGGCTCTTTA 240

Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80

QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCTCAGATCAGTTATTCCTAGTCTGACCTCA 300

Db 81 GlyThrTrpTrpProAspLeuTyValCysLeuArgSerValIleProSerLeuAlaThr 100

QY 301 CCCCCAGATATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACCCACCAATAATGGA 360

Db 101 ProProAspIleLeuHisAlaHisGlyPheTyValCysProGlyProProAsnAsnGlu 120

QY 361 AAACATTGCCGAATCCAGAGATTTCTTTTGTAAACAATGAACTGTGTAACTCTAAT 420

Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140

QY 421 GATGGATATTGGAAATGGCCAACTCTCAGCAGGATAGGGTAAGTTTCTTATGTCAAC 480

Db 141 AspGlyTyTrpLysTrpProThrSerGlnAspArgValSerPheSerTyValAsn 160

QY 481 ACCTATACCACTCTGGACAAATTTAATTAATCTGACCTGGATTAGAACTGGNAGCCCCAG 540

Db 161 ThrTyThrSerSerGlyGlnPheAsnTyLeuThrTrpIleArgThrGlySerProLys 180

QY 541 TGCTCTCTCTTCCAGACCTTAGATTACCTAAAAATAAGTTTCTACTGAGAAGCAAAACAAG 600

Db 181 CysSerProSerAspLeuAspTyLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200

QY 601 AATATCTAAAAATGGGTAAATGGTATGCTCTTGGGAATGGTATATTATGAGGCTCGGT 660



Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTy-TyrGlySerGly 220  
QY 661 AAACACAGGCTCCATTCTAACTATTTCGCTCAAAATAAACACGCTGGAGCTTCCAATG 720  
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
QY 721 GCTATAGACCAAAATACGGCTTGTAGCGGTCAAGACCCCAACCAAGGACCCAGGACCA 780  
Db 241 AlaThrGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
QY 781 TCCTCTAAACAACCTTCCTGGATCAGACCCCACTAGAGTCTAAACAGCACCACTAAATAATGGG 840  
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
QY 841 GCAAACTTTTATAGCTCATCCAGGAGCTTTTCAAGCTCTTAAGCTCTTAACCTCACGACTCCAGAG 900  
Db 281 AlalysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 901 GCTACCTCTTCTGTGTGGCTATGCTTAGCTTCGGGCCCACTTACTATGAAGGAATGGCT 960  
Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyTrpGluGlyMetAla 320  
QY 961 AGAAGAGGAATTCATGTGCAAAAGAACAATAGAGACCAATGCACATGGGGATCCCAA 1020  
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 1021 AATAAGCTTACCTTACTGAGTTTCTCGAAGGACACCTGCATAGGAAGGTTCCTCCCA 1080  
Db 341 AsnLysLeuThrLeuThrGluValPheGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 1081 TCCCAACCAACCTTTGTAAACACACTGAAGCTTTAATCAAACTCTGAGAGTCAATAT 1140  
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTy 380  
QY 1141 CTGTGTACCTGGTTATGACAGGTGGTGGCGATGATTAATCTGGAATTAACCCCTTGTTC 1200  
Db 381 LeuValProGlyTyTrpAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 1201 ACCTTGGTTTAAACAACTAAAGATTTTGCATTATGGTCCAAATTTGTTCCCGAGTG 1260  
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 1261 TATTACTATCCGAAAAGCAATCTCTGATGAATATGACTACAGAAATCATCGACAAAAG 1320  
Db 421 TyTrpTyTrpProGluLysAlaIleLeuAspGluTyTrpAspTyTrpGlyAsnHisArgGlnLys 440  
QY 1321 AGAGAACCCATATCTCTGACACTTGCTGTGATGCTCGGACTTGGAGTGGCAGCAGTGTA 1380  
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
QY 1381 GGAACAGGAACAGCTGCCCTGTCTACGGGACACAGCAGCTTAGAAACAGGACTTAGTAAC 1440  
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 1441 CTACATCGAATTTGAACAGACACTCTCCAAGCCCTAGAAATCTGTCAGTAACCTCGAG 1500  
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 1501 GAATCCCTAAACCTCTTATCTCAAGTAGTCTTACAGAATAGAGAAGGTTAGATTATTA 1560  
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
QY 1561 TTTCTAAAAGAGGAGTATGTGTAGCTTGAAGGAGGAATGCTGTTTTTATGTGGAT 1620  
Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysPheTyValAsp 540  
QY 1621 CATTCAGGGCCATCAGAGACTCCATGACAGCTTAGAGAAGGTTCGAGAAGCGTCA 1680  
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArg 560  
QY 1681 AGGGAAAAGGAACACTACTCAAGGGTGGTTTGAGGGATGTTCAACAGGTCTCTTTGGTTG 1740  
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580

QY 1741 GCTACCTTACTTTCTGCTTTAAACAGGACCCCTTAATAGTCTCTCTCTTACTCAGTT 1800  
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
QY 1801 GGGCCATGTATTATTAAACAAGTTAATTGCTTTCATTAGAGAACGAATAGTGCACTCCAG 1860  
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 1861 ATCATGTGTACTAGACACAGTACAAAGCCCGTCTAGCAGGGAAGCTGGCCGC 1914  
Db 621 IleMetValLeuArgGlnGlnTyTrpGlnSerProSerArgGluAlaGlyArg 638  
RESULT 7  
QY Q06Y9\_9GAMR PRELIMINARY; PRT; 641 AA.  
ID Q06Y9\_9GAMR PRELIMINARY; PRT; 641 AA.  
AC Q06Y9;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Envelope glycoprotein.  
GN Name=env;  
OS Porcine endogenous retrovirus.  
OC Viruses; Retrovirdae; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21851116; PubMed=11861871;  
RX DOI=10.1128/JVI.76.6.3045-3048.2002;  
RA Oldmixon B.A., Wood J.C., Ericsson T.A., Wilson C.A.,  
RA White-Scharf M.E., Andersson G., Greenstein J.L., Schuurman H.J.,  
RA Patience C.;  
RT "Porcine endogenous retrovirus transmission characteristics of an  
RT inbred herd of miniature swine."  
RL J. Virol. 76:3045-3048(2002).  
DR EMBL; AF417229; AAL87624.1; -; Genomic\_DNA.  
DR HSSP; P03385; 1WOF  
DR SMR; Q06Y9; 491-543.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002050; Env polyprotein.  
DR InterPro; IPR008981; FwuVrecept-bind.  
DR Pfam; PF04429; ILV\_coat; 1.  
KW Envelope protein.  
SQ SEQUENCE 641 AA; 71388 MW; 9CD82521A814923D CRC64;  
Alignment Scores:  
Pred. No.: 5,07e-258 Length: 641  
Score: 3293.50 Matches: 614  
Percent Similarity: 97.3% Conservative: 10  
Best Local Similarity: 95.8% Mismatches: 14  
Query Match: 95.0% Indels: 3  
Gaps: 1  
US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x Q06Y9\_9GAMR (1-641)  
QY 1 ATGCATCCACGCTTAAACCGCGCCACTCCCGATTCCGGGTGGAAAGCCGAAAGACTG 60  
Db 1 MethisProThrLeuSerArgHisLeuProfileArgGlyGlyLysProLysArgLeu 20  
QY 61 AAAATCCCTTAAGCTTCGCTCCATCGGTGTTCTTACTCTCTCAATACCTCTAG 120  
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40  
QY 121 ACTAATGTGTATGCGCATAGGAGACGCTGAACCTCCATAAACCTTATCTCTCCTGG 180  
Db 41 VallanGlyLysArgLeuValAsnSerProAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 181 TTAATTACTGACTCCGGCACAGGTATTAAATATCAACAACTCAAGGGAGGCTCTTTA 240  
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80



```
QY 241 GGAACCTGGTGGCTGATCTATAGCTTTGGCTCAGATCAGTTATTCTTAGTCTG----- 294
Db 81 GlyThrTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY 295 ---ACCTCACCCAGATATCTCATCTCCTCAGGATTTTATGTTGGCCAGGACCA 351
Db 101 GlnAlaThrProProAspValIleLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
QY 352 AATAATGGAACAACTTGGGAAATCCAGAGATTCTTTTAAACAATGAACCTGTGTA 411
Db 121 AsnAsnGluGluHisCysGlyAsnProArgaspPhePheCysLysGlnTrpAsnCysVal 140
QY 412 ACCTCTAATGATGATATGGAATGGCCAACTCTCAGCAGGATAGGGTAAAGTTTCT 471
Db 141 ThrSerAsnAspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSer 160
QY 472 TATGTCAACACCTATACAGCTCGGACAATTTAATACCTGACCTGGATTAGAACTGGA 531
Db 161 TyrValAsnThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGly 180
QY 532 AGCCCCAAGTCTCTCTTCAGACTAGATTACCTTAAATAAAGTTTCACTGAGAAAGGA 591
Db 181 SerProLysCysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGly 200
QY 592 AAACAAGAAATATCTTAAATGGGTAAATGGTATGTTCTGGGGAATGGTATATTATGGA 651
Db 201 LysGlnGluAsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGly 220
QY 652 GGCTCGGTAACAACCGGCTCATCTTAATCTTAATCTTAACTTAACTTAACTTAACTGAG 711
Db 221 GlySerGlyLysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGlu 240
QY 712 CCTCAATGGCTATAGGACCAATACGCTCTTGACGGGTCAAGACCCCAACCCCAAGGA 771
Db 241 ProPrometAlaThrGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGly 260
QY 772 CCAGGACCATCTCTAACAATCTCTGGATCAGACCCCACTGAGTCTTAACAGCAGCACT 831
Db 261 ProGlyProSerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThr 280
QY 832 AAAATGGGCGCAAACTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCAACG 891
Db 281 LysMetGlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThr 300
QY 892 ACTCCAGAGCTACTCTCTTGTGGCTATGCTTAGCTTCGGGCCCACTTACTATGAA 951
Db 301 ThrProGluAlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGlu 320
QY 952 GGAATGGCTAGAAGAGGAAATTCATATGTGACAAAGAACAATAGACCAATGCACATGG 1011
Db 321 GlyMetAlaArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrp 340
QY 1012 GGATCCCAAAATAGCTTACCTTACTGAGGTTTCTGGAAGGACCTGATAGGAAG 1071
Db 341 GlySerGlnAsnLysLeuThrLeuThrGluValPheGlyLysGlyThrCysIleGlyLys 360
QY 1072 GTTCCCCCATCCCAACCACTTTGTAAACCACTGACGCTTAACTTAACTTAACTCTGAG 1131
Db 361 ValProProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGlu 380
QY 1132 AGTCAATATCTGGTACCTGGTTATGACAGGTGGGCGCATGTAATCTGATTAAACCCCT 1191
Db 381 SerGlnTyrLeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrPro 400
QY 1192 TGTGTTTCCACTGGTGGTTTTTAAACCAATTAAGATTTTGCATTTATGGTCCAAATGTT 1251
Db 401 CysValSerThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleVal 420
QY 1252 CCCGAGGTATTACTATCCGAAAAGCAATCTTGCATGAATATGACTACAGAATCAT 1311
Db 421 ProArgValTyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrGlyAsnHis 440
```

```
QY 1312 CGACAAAGAGAGAACCCATATCTCTGACACTTGTGTGATGCTCGGACTTGGAGTGCA 1371
Db 441 ArgGlnLysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyValAla 460
QY 1372 GCAGGTGTAGAAACAGGAACAGCTGCCCTGTGTCTACGGGACCACACAGCAGCTGAAACAGA 1431
Db 461 AlaGlyValGlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGly 480
QY 1432 CTTAGTAACTTACATCGAATTTGTAAACAGAGATCTCCAAGCCCTAGAAAAATCTGTCTAGT 1491
Db 481 LeuSerAsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSer 500
QY 1492 AACCTGAGGAATCCCTAACCTCTCTTATCGAAGTAGTCTCTACAGAATAGAAAGGGTTA 1551
Db 501 AsnLeuGluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeu 520
QY 1552 GATTATATTATTTCTAAAGAGGAGGATTATGCTAGCTTGAAGGAGGAATGCTGTTTT 1611
Db 521 AspLeuLeuPheLeuLysGluGlyLeuCysValAlaLeuLysGluGluCysCysPhe 540
QY 1612 TATGTGATCATTCAGGGCCATCAGAGACTCCATGAACAGCTTAGAGAAAGGTTGGAG 1671
Db 541 TyrValAspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGlu 560
QY 1672 AAGCTCGAAGGGAAAGAACTACTCAAGGGTGGTTTGAAGGATGGTTCAACAGGTCT 1731
Db 561 LysArgArgArgGluLysGluThrThrGlnGlyTyrPheGluGlyTyrPheAsnArgSer 580
QY 1732 CTTTGGTGGCTACCTTCTTCTTAAACAGGACCTTAAATAGTCTCTCTCTGTTA 1791
Db 581 ProTrpLeuAlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeu 600
QY 1792 CTCACAGTGGCCCATGTATTATTAAACAGTAAATTCGCTTCAATAGAACGAAATAGT 1851
Db 601 LeuThrValGlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSer 620
QY 1852 GCAGTCCAGATCATGTGTTACTTAGACACAGTACCAAGCCCGCTAGCAGGGAAGCTGGC 1911
Db 621 AlaValGlnIleMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGly 640
QY 1912 CGC 1914
Db 641 Arg 641
RESULT 8
Q5Q0Q7_9GAMR PRELIMINARY; PRT; 653 AA.
AC Q5Q0Q7;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Env protein.
OC Porcine endogenous retrovirus C/A.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus;
OC 1-Mammalian type C virus group.
OX NCBI_TaxID=286072;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Al4/220;
RX PubMed=15564496; DOI=10.1128/JVI.78.24.13880-13890.2004;
RA Bartosch B., Stefanidis D., Myers R., Weiss R., Patience C.,
RA Takeuchi Y.;
RT "Evidence and consequence of porcine endogenous retrovirus
RT recombination.";
RL J. Virol. 78:13880-13890(2004).
DR EMBL; AY570980; AAT77168.1; -; Genomic_DNA.
DR SMR; Q5Q0Q7; 503-555.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR02050; Env_polyprotein.
DR Pfam; PF00429; TLV_coat; 1.
SQ SEQUENCE 653 AA; 73070 MW; 18BB5A80343B7511 CRC64;
```

## Alignment Scores:

Pred. No.: 1,15e-246 Length: 653  
Score: 3153.50 Matches: 594  
Percent Similarity: 93.6% Conservative: 17  
Best Local Similarity: 91.0% Mismatches: 27  
Query Match: 90.9% Indels: 15  
DB: 2 Gaps: 3

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x Q50GQ7\_9GNMR (1-653)

```
QY 1 ATGCATCCAGCTTAAACGGCGCCACCTCCGATTCGGGTGGAAAGCCGAAAGACTG 60
D 1 MethHisProThrLeuSerArgArgHisLeuProIleArgGlyGlyLeuProLysArgLeu 20
QY 61 AAATCCCTTAAGCTTCGCTCCATCGGTGGTTCCTTACTCTGTCAATTAACCTCTCAG 120
D 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
QY 121 ACTAATGCTATCGCATAGGACAGACCTGAACTCCCATAAACCTTATCTCTCACCTGG 180
D 41 ValAsnGlyLysArgLeuValAsnSerProAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTACTGATCCGGCACAGGTATTAAATATCAACAACTCAAGGGAGGCTCTTTA 240
D 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTCAGATCAGTTATTCTTAGTCTG----- 294
D 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY 295 ---ACCTCACCCAGATATCTCCATGCTCAGCGATTTTATGTTGCCAGGACCCACA 351
D 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
QY 352 AATAATGAAACATTCGGGAAATCCAGAGATTTCTTTGTAAACAATGGAACCTGTGTA 411
D 121 AsnAsnGluGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
QY 412 ACCTCTAATGATGGATATTTGGAATGGCAACCTCTCAGCAGGATAGGGTAAGTTTCT 471
D 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160
QY 472 TATGTCAACACCTATACACGCTCTCGACAATTTAATTAC-----CTGACC 516
D 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
QY 517 TGG-----ATTAGAACTGGAAGCCCCCAAGTGTCTCTCTTCAGAC 555
D 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
QY 556 CTAGATTACCTMAAATAAGTTTCACTGAGAAGGAAACAGAAATAATCTTAAATGG 615
D 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleLeuLysTrp 220
QY 616 TTAATGATGATCTTGGGGAATGCTATATATGAGGCTGGGTAAACAACACGAGCTCC 675
D 221 ValAsnGlyMetSerTrpGlyMetValTyrTyrGlySerGlyLysGlnProGlySer 240
QY 676 ATTCTAACTATTTCGCTCAAAATAAACACGAGCTGGAGCCTCCAAATGGCTATAGGACCAAT 735
D 241 IleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMetAlaIleGlyProAsn 260
QY 736 ACGGTCTGACGGGTCAAGACCCCAACCCAGGACGAGGACCATCTCTTAACATAACT 795
D 261 ThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyProSerSerAsnIleThr 280
QY 796 TCTGATCAGACCCCACTGAGTCTAACAGCAGCTAAATAATGGGGCAAACTTTTATAGC 855
D 281 SerGlySerAspProThrGluSerAsnSerThrThrLysMetGlyAlaLysLeuPheSer 300
QY 856 CTCATCCAGGGAGCTTTTCAAGCTCTTAATCTCCAGCTCCAGAGGCTACCTTTCTTGT 915
D 301 LeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThrSerSerCys 320
```

## RESULT 9

Q5J1S8\_SUSBA  
ID Q5J1S8\_SUSBA PRELIMINARY; PRT; 642 AA.  
AC Q5J1S8;

```
QY 916 TGGCTATGCTTAGCTTCGGGCCACCTTACTATGAAGGAATGGCTAGAGAGGGAATTC 975
D 321 TrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAlaArgArgGlyLysPhe 340
QY 976 AATGTGACAAAGAACATAGACACCAATGCAATGGGATCCCAAAATAAGCTTACCCCTT 1035
D 341 AsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGlnAsnLysLeuThrLeu 360
QY 1036 ACTGAGGTTTCTGAAAAGGCACCTGCTAGTAAGAAAGTTTCCCCCATCCCAACACCTT 1095
D 361 ThrGluValSerGlyLysGlyThrCysIleGlyLysValProProSerHisGlnHisLeu 380
QY 1096 TGTAAACACACTGAAGCCCTTTAATCAAACTCTCAGAGTCAATATCTGGTACCTGGTTAT 1155
D 381 CysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyrLeuValProGlyTyr 400
QY 1156 GACAGGTGGTGGGCATGTAATACTAGGATTAACCCCTTGTGTTCACCTTGGTTTTTAAC 1215
D 401 AspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeuValPheAsn 420
QY 1216 CAAACTAAGATTTTTCGATTTGTCATGCTACAGAAATCATCGACAAAGAGAGAACCATATCT 1275
D 421 GlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrTyrProGlu 440
QY 1276 AAAGCAATCCTTGTATGATGATGCTACAGAAATCATCGACAAAGAGAGAACCATATCT 1335
D 441 LysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLysArgGluProLysSer 460
QY 1336 CTGACACTTGTGTGTGCTCGGACTTGGAGTGGCAGCAGGTGTAGGAAACAGGACAGCT 1395
D 461 LeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyValGlyThrGlyThrAla 480
QY 1396 GCCCTGGTCAAGGACACACGACGCTAGAAACAGGACTTAGTAACCTACATCGAATTGTA 1455
D 481 AlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHisArgIleVal 500
QY 1456 ACAGAAGATCTCCAAGCCCTAGAAAATCTCTCAGTAACCTGGAGGAATCCTTAACCTCC 1515
D 501 ThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluLysSerLeuThrSer 520
QY 1516 TTATCTGAAGTAGTCTCTACAGAATAGAGGGTTAGATTATTATTCTTAAAGAAGGA 1575
D 521 LeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeuPheLysGluGly 540
QY 1576 GGATTATGCTAGCTTGAAGGAGATGCTGTTTTTATGTTATGTTATGTTATGTTATGTTATG 1635
D 541 GlyLeuCysValAlaLeuLysGluCysCysPheTyrValAspHisSerGlyAlaIle 560
QY 1636 AGAGACTCATGAACAAGCTTTAGAGAAAGTTGGAGAGAGCTCGAAGGAAAGGAAAGAACT 1695
D 561 ArgAspSerMetSerLysLeuArgGluArgLeuGluLysArgArgArgGluLysGluThr 580
QY 1696 ACTCAAGGGTGGTTTGGGGATGGTTCAACAGGTCTCTTTGGTTGGCTACCTACTTTCT 1755
D 581 ThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeuAlaThrLeuLeuSer 600
QY 1756 GCTTTAACAGGACCTTAAATAGTCTCTCTCTGTTACTCACAGTTGGGCCCATGTTATT 1815
D 601 AlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrValGlyProCysIleIle 620
QY 1816 AACAGTTTAATTTGCTTCATTAGAGAACGAATAAGTGCAGTCCAGATCATGTACTTAGA 1875
D 621 AsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMetValLeuArg 640
QY 1876 CAACAGTACCAAGCCCGCTCTAGCAGGGAAGCTGGCCGC 1914
D 641 GlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 653
```

DT 10-MAY-2005 (TrEMBLrel. 30, Created)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
 DE Envelope glycoprotein.  
 GN Name=env;  
 OS Sus barbatus (Bearded pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
 OC Sus.  
 OX NCBI\_TaxID=41807;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=1559862; DOI=10.1128/JVI.79.1.649-654.2005;  
 RA Niebert M., Tonjes R.R.;  
 RT "Evolutionary spread and recombination of porcine endogenous  
 RT retroviruses in the suiformes.";  
 RL J. Virol. 79:649-654 (2005).  
 DR EMBL: AY534305; AAT06257.1; -; mRNA.  
 DR GO: GO:0019028; C:viral capsid; IEA.  
 DR GO: GO:0019031; C:viral envelope; IEA.  
 DR GO: GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro: IPR002050; Env polyprotein.  
 DR Pfam: PF00429; TIV\_coat; 1.  
 KW Envelope protein.  
 SQ SEQUENCE 642 AA; 71667 MW; 2B5EAD8A4DFBE60C CRC64;

Alignment Scores:  
 Pred. No.: 2,2e-246 Length: 642  
 Score: 3150.00 Matches: 597  
 Percent Similarity: 95.5% Conservative: 17  
 Best Local Similarity: 92.8% Mismatches: 24  
 Query Match: 90.8% Indels: 6  
 DB: 2 Gaps: 2

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x Q5J1S8\_SUSBA (1-642)

Qy	1	ATGCATCCCAAGTTAAACCGCGGCACCTCCCGATTGGGTGGAAAGCCGAAAGACTG	60
Db	1	MethHisProThrLeuSerArgHisLeuProIleArgGlyGlyLeuProLysArgLeu	20
Qy	61	AAATACCCCTTAAGCTTCCCTCCATCGCGTGGTTCCTTACTCTGTCATTAACCTCTCAG	120
Db	21	LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln	40
Qy	121	ACTAATGTATGCGCATAGAGACAGACCTGAACTCCCAATAACCTTATCTCTCACCTGG	180
Db	41	ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp	60
Qy	181	TTAATTACTGACTCCGGCACAGGTATTAAATATCAACAACACTCAAGGGAGGCTCTTTA	240
Db	61	LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu	80
Qy	241	GGRACTGGTGGCTGATCTATAGTTGGTTCCTCAGATCAGTTATTCCTAGTCTG-----	294
Db	81	GlyThrTrpTrpProGluLeuTyValCysLeuArgSerValIleProGlyLeuAsnAsp	100
Qy	295	---ACCTCACCCACAGATATCTCCATGCTCAGCGATTTATGTTTCCCGAGGACCA	351
Db	101	GlnThrSerProProAspIleLeuArgAlaHisGlyPheTyValCysProGlyProPro	120
Qy	352	AATTAATGGAACAATTCGGAAATCCCGAGATTTCTTTGTAAACAATGGAATGTGTA	411
Db	121	AsnAsnGluLysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysVal	140
Qy	412	ACCTCTAATGATGATATGGAATGGCCACCTCTCAGCAGGATAGGTAAGTTTCT	471
Db	141	ThrSerAsnAspGlyTyTrpLysTrpProThrSerGlnGlnAspArgValSerPheSer	160
Qy	472	TATGTCACACCTATACAGCTCTGGACAATTTAATTACCTTGACCTGGATTAGAACTGGA	531
Db	161	TyrValAsnThrTyThrSerSerGlyGlnPheAsnTyLeuThrTrpIleArgThrGly	180
Qy	532	AGCCCCAAGTGCTCTCTTCAGACCTAGATTACCTAAATAAAGTTTCTACTGAGAAAGGA	591

Db	181	SerProLysCysSerProSerAspLeuAspTyLeuLysIleSerPheThrGluLysGly	200
Qy	592	AAACAAGAAATATCCTAAATGGTAAATGGTATGCTTGGGGAATGGTATATATGGA	651
Db	201	LysGlnGluAsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyTrpGly	220
Qy	652	GGCTCGGGTAAACAACAGGCTCCATCTTAATCTTAACTATTCGCCTCAAAATAAACCA-CCTCGA	710
Db	221	GlySerGlyLysLysProGlySerIleLeuThrIleArgLeuLysIleAsnSerAlaGly	240
Qy	711	GCCTCCAAATGGCTATAGACCAAAATACGGTCTTCAACGGGTCAAGACCCCAACCAAGG	770
Db	241	AlaSerAsnCysTyArgThrLysTyGlyLeuAspGlySerArgThrSerAsnProArg	260
Qy	771	ACGAGGACCATC---CTCTACATAACTCTGGATCAGACCCCACTGATCTACAGCAC	827
Db	261	ThrSerIleTyProLeu-AsnThrThrSerGlySerValProThrGluSerAsnSerTh	280
Qy	828	GACTAAATGGGGCAAAACCTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACCTC	887
Db	280	rThrLysMetGlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSe	300
Qy	888	CACGACTCCAGAGGCTACCTCTTCTTGTGGCTATGCTTAGCTTCGGGCCCACTTACTA	947
Db	300	rThrThrProGluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyTy	320
Qy	948	TGAAGGAATGGCTAGAGAGGAATTCATGTGCACAAAGAACATAGAGACCAATGCAC	1007
Db	320	rGluGlyMetAlaArgGlyGlyLysPheAsnValThrLysGluHisArgAspGlnCysTh	340
Qy	1008	ATGGGATCCCAAAATAAGCTTACCTTACTAGGTTTCTGAAAAGCACCTGCATAGG	1067
Db	340	rTrpGlySerGlnAsnLysLeuThrLeuValSerGlyLysGlyThrCysIleGln	360
Qy	1068	AAAGTTTCCCCCATCCCAACAACCTTTGTAAACCACTGAAGACCTTTTAAATCAAACTC	1127
Db	360	yMetValProProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSe	380
Qy	1128	TGAGAGTCAATACTCGTACTCGTTATGACNGTGGTGGCATGTAACTATGATTAAC	1187
Db	380	rGluSerGlnTyLeuValProGlyTyAspArgTrpTrpAlaCysAsnThrGlyLeuTh	400
Qy	1188	CCCTTGTTGTTCCACCTTGGTTTAAACCAAACTTAAAGATTTTGTCAATATGTCCTCAAT	1247
Db	400	rProCysValSerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIl	420
Qy	1248	TGTTTCCCGAGTGATTTACTATCCGAAAAGCAATCCTTGATGAATATGACTACAGAAA	1307
Db	420	eValProArgValTyTyTrpProGluLysAlaValLeuAspGluTyAspTyArgTy	440
Qy	1308	TCATCGACAAAGAGAGAACCCATATCTCTGACACTTGTGTGTGATGCTCGGACTTGGAGT	1367
Db	440	rAsnArgProLysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyVa	460
Qy	1368	GGCAGCAGGTGTAGGAACAGAACAGCTGCCCTGGTCACGGGACACAGCAGCTAGAAAC	1427
Db	460	AlaAlaGlyValGlyThrGlyThrAlaLeuIleThrGlyProGlnGlnLeuGluLy	480
Qy	1428	AGGACTTGTAGTAACTACATCGAATTGTAAACAGAGATCTCCAAGCCCTAGAAAAATCTGT	1487
Db	480	sGlyLeuSerAsnLeuHisArgIleValThrGluAsnLeuGlnAlaLeuGluLysSerVa	500
Qy	1488	CAGTACCTCGGGAATCCCTTAACCTCTTATCTGAGTAGTCTCTACAGATAGAGNGG	1547
Db	500	lSerAsnLeuGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGl	520
Qy	1548	GTTAGATTATTTCTTAAAGAGGAGGATTATGTGTAGCTTGAAGCAGGAGTAATGCTG	1607
Db	520	yLeuAspLeuLeuPheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCy	540
Qy	1608	TTTTTATGTGGATCAITTCAGGGGCCATCAGAGACTCCATGAACAGCTTAGAGAAAGGTT	1667

```

Db      540  sPheTyRValAspHisSerGlyAlaIleAArgAspSerMetSerLysLeuAArgGluArgLe 560
QY      1668  GGAGAGCGTGAAGGGAAGAACTACTCAAGCGTGGTTGAGGATGGTTCAACAG 1727
Db      560  uGluAArgArgArgGluAArgGluAAspGlnGlyTrpPheGluGlyTrpPheAsnAr 580
QY      1728  GTCTCTTTGGTTGGCTACCTACTTTCGTCTTAAACAGACCCCTTAATAGTCCTCCTCT 1787
Db      580  gSerProTrpMetAlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeu 600
QY      1788  GTTACTCACAGTTGGCCCATGTATTATTAAACAAGTTAATTCCTTCATTAGAGAACGAAT 1847
Db      600  uLeuLeuThrValGlyProCysAlleIleAsnLysLeuIleAlaPheIleAArgGluArgI 620
QY      1848  AAGTGCAGTCAGATCATGTACTTAGACACAGTACCAGAGCCCGTCTAGCAGGGAAGC 1907
Db      620  eSerAlaValGlnIleMetValLeuAArgGlnGlnTyRglnSerProSerSerAArgGluAl 640
QY      1908  TGGCGCG 1914
Db      640  aGlyArg 642

RESULT 10
OQ0620_9GAMR PRELIMINARY; PRT; 653 AA.
AC OQ0620_
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope glycoprotein.
GN Name=env;
OS Porcine endogenous retrovirus.
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus;
OC 1-Mammalian type C virus group.
OX NCBI TaxID=61673;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21851116; PubMed=11861871;
RX DOI=10.1128/JVI.76.6.3045-3048.2002;
RA Oldmixon B.A., Wood J.C., Ericsson T.A., Wilson C.A.,
RA White-Schaff M.E., Andersson G., Greenstein J.L., Schuurman H.J.,
RA Patience C.;
RT "Porcine endogenous retrovirus transmission characteristics of an
RT inbred herd of miniature swine.";
RL J. Virol. 76:3045-3048(2002).
DR EMBL; AF417228; AAL87623.1; -; Genomic_DNA.
DR HSSP; P03385; IMOF.
DR SMR; OQ0620; 503-555.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env polyprotein.
DR InterPro; IPR008981; FMuLVrecept-bind.
DR Pfam; PF00429; TLV_coat; 1.
KW Envelope protein.
SQ SEQUENCE 653 AA; 73113 MW; 4A5959F988B9B3C CRC64;

Alignment Scores:
Pred. No.: 3.53e-246 Length: 653
Score: 3147.50 Matches: 593
Percent Similarity: 93.6% Conservative: 18
Best Local Similarity: 90.8% Mismatches: 27
Query Match: 90.8% Indels: 15
DB: 2 Gaps: 3

US-10-723-552-3_COPY_5620_7533 (1-1914) x OQ0620_9GAMR (1-653)
QY 1 ATGCATCCAGCTTAACCGCGCCACCTCCGATTCCGGGTGGGAAGCCGAAGACTG 60
Db 1 MethHisProThrLeuSerAArgHisLeuProIleAArgGlyLysProIleAArgLeu 20
QY 61 AAAATCCCTTAAGCTTCGCTCCATCGGTGGTTCCCTTACTCTGTCAATAACCTCTCAG 120

```

```

Db      21  LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
QY      121  ACTAATGGTATGCGATAGAGACAGCTGAACCTCCATAAACCCTTATCTCTCACCTGG 180
Db      41  ValAsnGlyLysAArgLeuValAsnSerProAsnSerHisLysProLeuSerLeuThrTrp 60
QY      181  TTAATTAAGTCTGCTCCGACACAGGTATTAAATATCAACAACTCAAGGAGGAGCTCTTTA 240
Db      61  LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAArgProLeu 80
QY      241  GGAACTGGTGGCTGTATATACGTTTGGCTTCAGATCAGTTATTCCTAGTCTG----- 294
Db      81  GlyThrTrpTrpProGluLeuTyRValCysLeuAArgSerValIleProGlyLeuAsnAsp 100
QY      295  ---ACCTCACCCCGAGATATCCTCATGCTCACCGATTTATGTTGGCCAGGACCACCA 351
Db      101  GlnAlaThrProProAspValLeuAArgAlaTyRglyPheTyRValCysProGlyProPro 120
QY      352  AATAATGGAACCAATTCCGGAATCCAGAGATTTCTTTTGTAAACAATGGAACCTGTGTA 411
Db      121  AsnAsnGluGluTyRCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
QY      412  ACCTTAATGATGATATGGAATGCCAACTCTCAGCAGGATAGGTTAAGTTTCT 471
Db      141  ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspAArgValSerTyRSer 160
QY      472  TATGTCAACACCTATACAGCTCTGGACATTTTAATTAC-----CTGACC 516
Db      161  PheValAsnAsnProThrSerTyRAsnGlnPheAsnTyRGlyHisGlyAArgTrpLysAsp 180
QY      517  TGG-----ATTAGAACTGGAAGCCCAAGTCTCTCTTCCAGAC 555
Db      181  TrpGlnGlnAArgValGlnLysAspValAArgAsnLysGlnIleSerCysHisSerLeuAsp 200
QY      556  CTAGATTACCTAAATAAAGTTTCACTGAGAAAGGAAACAAAGAAATATCTCTAAATGG 615
Db      201  LeuAspTyRLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleLeuLysTrp 220
QY      616  GTAATGATGATGCTCTGGGNAATGGTATATTATGAGGCTCGGCTAAACACCAAGCTCC 675
Db      221  ValHisGlyMetSerTrpGlyMetValTyRTrpGlyGlySerGlyLysGlnProGlySer 240
QY      676  ATTCTAACTATTTCGCTCAAAATAAAACAGCTGAGGCTCCCAATGGCTATAGGACCAAT 735
Db      241  IleLeuThrIleAArgLeuLysIleAsnGlnLeuGluProProMetAlaIleGlyProAsn 260
QY      736  ACGGTCTTGACGGGTCAAAAGACCCCAACCCAGGACAGGACCATCTCTTAAACATAACT 795
Db      261  ThrValLeuThrGlyGlnAArgProProThrGlnGlyProGlyThrSerSerAsnIleThr 280
QY      796  TCTGATCAGACCCCACTGAGTCTAACAGCAGCTAAATGGGGGCGGCGGCGGCGGCGGCG 855
Db      281  SerGlySerAspProThrGluSerAsnSerThrThrLysMetGlyAlaLysLeuPheSer 300
QY      856  CTCTATCCAGGAGCTTTTCAAGCTCTTAACCTCCACGACTCCAGAGGCTACCTCTCTTGT 915
Db      301  LeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThrSerSerCys 320
QY      916  TGGCTATGCTTGTTCGGGCCCACTTACTATGAAGAATGGCTAGAGAGGGAATTC 975
Db      321  TrpLeuCysLeuAlaLeuGlyProProTyRTrpGluGlyMetAlaAArgAArgGlyLysPhe 340
QY      976  AATGTGCAAAAGAACATAGACCAATGCATGGGGATCCCAAAATAAGCTTACCCCTT 1035
Db      341  AsnValThrLysGluHisAArgAspGlnCysThrTrpGlySerGlnAsnLysLeuThrLeu 360
QY      1036  ACTGAGGTTTCTGGAAAGGCACCTGCATAGAAAGGTTCCCCCATCCCAACCAACCTT 1095
Db      361  ThrGluValSerGlyLysGlyThrCysIleGlyLysValProProSerHisGlnHisLeu 380
QY      1096  TGTAAACCACTGAAGCCTTTAATCAAACTCTCAGAGTCAATATCTGGTACCTGGTTAT 1155
Db      381  CysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyRLeuValProGlyTyR 400

```

QY 1156 GACAGGTGGTGGGATGTAATACCTGATTAACCCCTTGTGTTCACCTGGTTTAAAC 1215  
DB 401 AspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeuValPheAsn 420  
QY 1216 CAACAAAGATTGTCATATGCTCCAAATGCTCCCGAGTGTTACTATCCCGAA 1275  
DB 421 GlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrTyrProGlu 440  
QY 1276 AAAGCAATCCTGTATGATATGACTACAGAAATCATCGACAAAAGAGAACCCATATCT 1335  
DB 441 LysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLysArgGluProIleSer 460  
QY 1336 CTGACACTGCTGTGATGCTCGACTTGGAGTGGCAGAGGTGTAGGAACAGAACAGCT 1395  
DB 461 LeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyValGlyThrGlyThrAla 480  
QY 1396 GCCCTGTCACGGACACACAGCAGCTAGAAACAGACTTAGTACCTACATCGAATTGTA 1455  
DB 481 AlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHisArgIleVal 500  
QY 1456 ACAGAAGATCTCCAAAGCCCTAGAAAATCTGTCACTAACCTGGAGGAATCCCTAACCTCC 1515  
DB 501 ThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluLysLeuThrSer 520  
QY 1516 TTATCTGAAGTAGCTCTACAGAAATAGAGGGTTAGATTATTATTTCTAAAGAAAGGA 1575  
DB 521 LeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeuPheLeuLysGluGly 540  
QY 1576 GGATATGTGTAGCTTGAAGGAGAAATCTGTTTATGTGGATCAATCGGGGCCATC 1635  
DB 541 GlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAspHisSerGlyAlaIle 560  
QY 1636 AGACATCCATGAACAGCTTAGAGAAAGGTTGAGAGAGCGTCGAAGCGAAAGGAACT 1695  
DB 561 ArgAspSerMetSerLysLeuArgGluArgLeuGluLysArgArgGluLysGluThr 580  
QY 1696 ACTCAAGGTGGTTGAGGAGTGTCAACAGGTCCTCTTTGGTGGCTACCTACTTTCT 1755  
DB 581 ThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeuAlaThrLeuLeuSer 600  
QY 1756 GCCTTAACAGGACCTTAATAGTCTCTCTCTGTACTACAGTTGGCCATGATTATT 1815  
DB 601 AlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrValGlyProCysIleIle 620  
QY 1816 AACAGTTAATTTGCTTTCATTAGAGAACGAATAGTCAGTCAGATCAGTACTAGTAGA 1875  
DB 621 AsnLysLeuIleAlaPheIleArgGluArgLysSerAlaValGlnIleMetValLeuArg 640  
QY 1876 CAACAGTACCAAGCCGCTCTAGCAGGAAGCTGGCCG 1914  
DB 641 GlnGlnTyrGlnSerProSerArgGluAlaGlyArg 653

## RESULT 11

ID Q0621\_9GAMR PRELIMINARY; PRT; 653 AA.  
AC Q0621;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Envelope glycoprotein.  
GN Name=env;

OS Porcine endogenous retrovirus.  
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus;  
OC 1-Mammalian type C virus group.  
OX NCBI\_TaxID=61673;  
RN [1]

## NUCLEOTIDE SEQUENCE.

RX MEDLINE=21851116; PubMed=11861871;  
RX DOI=10.1128/JVI.76.6.3045-3048.2002;  
RA Oldmixon B.A., Wood J.C., Ericsson T.A., Wilson C.A.,  
RA White-Scharf M.E., Andersson G., Greenstein J.L., Schuurman H.J.,  
RA Patience C.;

RT "Porcine endogenous retrovirus transmission characteristics of an  
RL inbred herd of miniature swine.";  
RL J. Virol. 76:3045-3048(2002).  
DR EMBL; AF417227; AAL87622.1; -; Genomic\_DNA.  
DR HSP; P01385; INOF.  
DR SMR; Q0621; 503-555.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002050; Env\_polyprotein.  
DR InterPro; IPR008981; FMuLVrecept-bind.  
DR Pfam; PF00429; Tlv\_coat; 1.  
KW Envelope protein.  
SQ SEQUENCE 653 AA; 73056 MW; 9E9D00137DDBA244 CRC64;

## Alignment Scores:

Pred. No.: 3,32e-245 Length: 653  
Score: 3135.50 Matches: 591  
Percent Similarity: 93.1% Conservative: 17  
Best Local Similarity: 90.5% Mismatches: 30  
Query Match: 90.4% Indels: 15  
DB: 2 Gaps: 3

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x Q0621\_9GAMR (1-653)

QY 1 ATGCATCCACGTTAAACCGCGCCACCTCCCGATTCGGGTGGAAGCCGAAAGACTG 60  
DB 1 MethHisProThrLeuSerArgHisLeuProIleArgGlyLysProLysArgLeu 20  
QY 61 AAAATCCCTTAAGTTCGCTCATCGCGTGTCTTACTCTGTCAATAACTCTCAG 120  
DB 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40  
QY 121 ACTAATGGTATGCGATAGGAGACAGCTCACTCCCATAAACCTTATCTCTACCTGG 180  
DB 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 181 TTAATTACTGACCTCCGCGCACAGGTATTAAATATCAACACACTCAAGGGGAGGCTCTTTA 240  
DB 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80  
QY 241 GGAACCTGTGGCTGATCTATAGTTTGGCTCAGATCAGTATTCTAGTCTG----- 294  
DB 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100  
QY 295 ---ACCTCACCCCGCATATCTCCATGCTCAGGATTTTATGTTGCCCGAGGACCA 351  
DB 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120  
QY 352 AATAATGGAACAATTCGCGAAATCCAGAGATTTCTTTGTAAACAATGGAACCTGTGTA 411  
DB 121 AsnAsnGluGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140  
QY 412 ACCTCTAATGATGATTTGGAATGCGCAACTCTCAGCAGGATAGGGTAAGTTTCT 471  
DB 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160  
QY 472 TATGTCACACCTATACCGCTCTGGACATTTAATTAC-----CTGACC 516  
DB 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180  
QY 517 TGG-----ATTAGAACTGGAAGCCCAAGTCTCTCTCTCTAGAC 555  
DB 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200  
QY 556 CTAGATTACCTAAAAATAGTTTCTACGTAGAGAGGAAACACAGAAAATATCTTAAATGG 615  
DB 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleLeuLysTrp 220  
QY 616 GTAAATGGTATGCTTGGGGAATGGTATATTATGGAGGCTCGGTTAAACACAGGCTCC 675  
DB 221 ValAsnGlyMetSerTrpGlyMetValTyrTyrGlySerGlyLysGlnProGlySer 240

```
Qy 676 ATTCTAACTATTTCGCTCAAAATAAACACGAGCTGGAGCTCCAAATGGCTATAGGACCAAAAT 735
Db |||||
Qy 241 ILeuThrIleArgLeuLysIleAsnGlnLeuGluProMetAlaIleGlyProAsn 260
Db |||||
Qy 736 ACGGTCTTTGACGGGTCAAAGACCCCAACCAAGACCAAGACCATCTCTTAAACATAACT 795
Db |||||
Qy 261 ThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyProSerSerAsnIleThr 280
Db |||||
Qy 796 TCTGGATCAGACCCCACTGAGCTTAAACAGACGACTAAATGGGGGCAAACTTTTACG 855
Db |||||
Qy 281 SerGlySerAspProThrGluSerAsnSerThrThrLysMetGlyAlaLysLeuPheSer 300
Db |||||
Qy 856 CTCATCCAGGAGCTTTTCAAGCTCTTAACTCCAGACTCCAGAGCTCCAGAGCTTCTTCTGT 915
Db |||||
Qy 301 LeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThrSerSerCys 320
Db |||||
Qy 916 TGGCTATGCTTAGCTTCGGGGCCACCTTACTATGAAGGAATGGCTAGAAAGGGAAATTC 975
Db |||||
Qy 321 TrpLeuCysLeuAlaLeuGlyProProThrTyrrGluGlyMetAlaArgArgGlyLysPhe 340
Db |||||
Qy 976 AATGTGACAAAGAACATAGACCAATGCACATGGGGATCCCAAAATAAGCTTACCTT 1035
Db |||||
Qy 341 AsnValThrLysGlyHisArgAspProCysThrTrpGlySerGlnAsnLysLeuThrLeu 360
Db |||||
Qy 1036 ACTGAGGTTTCTGGAAGGACCTGTCATAGAAAGGTTCCCCATCCCAACACACCTT 1095
Db |||||
Qy 361 ThrGluValPheGlyLysGlyThrCysIleGlyLysValProProSerHisGlnHisLeu 380
Db |||||
Qy 1096 TGTAAACCACTGAAGCCCTTAAATCAAACTCTGAGAGTCAATATCTGTGTAACCTGGTTAT 1155
Db |||||
Qy 381 CysAsnHisThrGluAlaPheAsnArgThrSerGluSerGlnTyrrLeuValProGlyTyrr 400
Db |||||
Qy 1156 GACAGGTGGTGGGCATGTAATACCTGAGTAACCCCTGTGTGTTCCACCTTGGTTTTTAAC 1215
Db |||||
Qy 401 AspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeuValPheAsn 420
Db |||||
Qy 1216 CAATAAAGATTTTGCATTATGTCCTCAAAATGTTCCCGAGTGTATTACTATCCCGAA 1275
Db |||||
Qy 421 GlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrrTyrrProGlu 440
Db |||||
Qy 1276 AAAGCAATCCTTGATGATATGACTACAGAAATCATCGACAAAAGAGAACCCATATCT 1335
Db |||||
Qy 441 LysAlaIleLeuAspGluTyrrAspTyrrArgAsnHisArgGlnLysArgGluProIleSer 460
Db |||||
Qy 1336 CTGACACTGCTGTGATCTCGACTTCGAGTGGCAGCAGGTGTAGGAACAGGACAGCT 1395
Db |||||
Qy 461 LeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIleValGlyValGlyThrAla 480
Db |||||
Qy 1396 GCCCTGGTCAGGGACACAGCAGCTAGAAACAGGACTTAGTAACCTACATCGAATTGTA 1455
Db |||||
Qy 481 AlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHisArgIleVal 500
Db |||||
Qy 1456 ACAGAAGATCTCCAGCCCTAGAAAAATCTGTCAAGTAACCTGGAGGAATCCCTAACCTCC 1515
Db |||||
Qy 501 ThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluSerLeuThrSer 520
Db |||||
Qy 1516 TTAATCTGAAGTAGTCTCAGAAATAGAGAGGGTTAGATTATTATTCTTAAAGAGGA 1575
Db |||||
Qy 521 LeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeuPheLeuLysGluGly 540
Db |||||
Qy 1576 GGATTATGTGTAGCTTCAAGGAGGATGCTGTTTTTTATGTGGATCATTCAGGGGCCATC 1635
Db |||||
Qy 541 GlyLeuCysValAlaLeuLysGluGluCysCysPheTyrrValAspHisSerGlyAlaIle 560
Db |||||
Qy 1636 AGAGACTCCATGAACAAGCTTAGAAAAAGGTTGAGAGAGCGCTCGAAGGGAAAAAGAACT 1695
Db |||||
Qy 561 ArgAspSerMetSerLysLeuArgGluArgLeuGluLysArgArgGluLysGluThr 580
Db |||||
Qy 1696 ACTCAAGGGTGGTTGAGGGATGGTTCAACAGGTCTCTTTGGTGGCTTACCCCTACTTCT 1755
Db |||||
Qy 581 ThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeuAlaThrLeuLeuSer 600
Db |||||
Qy 1756 GCTTTAACAGGACCCCTTAATAGTCTCTCTCTGTTACTCACAGTGGGGCCATGATTATT 1815
Db |||||
```

```
Db 601 AlalaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrValGlyProCysIleIle 620
Qy 1816 AACAGTAATATGCTTCATTAGAGAACGATAAAGTCGAGTCAGATCATGGTACTTAGA 1875
Db |||||
Qy 621 AsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMetValLeuArg 640
Db |||||
Qy 1876 CMACAGTACCAAGCCGCTCTAGCAGGGAAGCTGGCCGC 1914
Db |||||
Qy 641 GlnGlnTyrrGlnSerProSerSerArgGluAlaGlyArg 653
Db |||||
RESULT 12
Q8Q6Y8_9GAMR
ID Q8Q6Y8_9GAMR PRELIMINARY; PRT; 659 AA.
AC Q8Q6Y8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Envelope glycoprotein.
GN Name=env;
OS Porcine endogenous retrovirus.
OC Viruses; Retrovirdae; Retroviridae; Gammaretrovirus;
OC 1-Mammalian type C virus group.
OX NCBI_TaxID=61673;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21851116; PubMed=11861871;
RX DOI=10.1128/JVI.76.6.3045-3048.2002;
RA Oldmixon B.A., Wood J.C., Ericsson T.A., Wilson C.A.,
RA White-Scharf M.E., Andersson G., Greenstein J.L., Schuurman H.J.,
RA Patience C.;
RT "Porcine endogenous retrovirus transmission characteristics of an
RT indred herd of miniature swine.";
RL J. Virol. 76:3045-3048(2002).
DR EMBL; AF417230; AAL87625.1; -; Genomic_DNA.
DR HSSP; P03385; 1MOF.
DR SMR; Q8Q6Y8; 509-561.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_Polyprotein.
DR InterPro; IPR008981; FmULVrecept-bind.
DR Pfam; PF00429; TLV_coat; 1.
KW Envelope protein.
SQ SEQUENCE 659 AA; 73902 MW; 1D0C1D84466B52F CRC64;
Alignment Scores:
Pred. No.: 1.02e-235 Length: 659
Score: 3018.50 Matches: 575
Percent Similarity: 90.4% Conservative: 21
Best Local Similarity: 87.3% Mismatches: 42
Query Match: 87.0% Indels: 21
DB: 2 Gaps: 6
US-10-723-552-3_copy_5620_7533 (1-1914) x Q8Q6Y8_9GAMR (1-659)
Qy 1 ATGCATCCAGCTTAACCGCGCACCTCCGATTCCGGGTGGGAACCGAAGACTG 60
Db |||||
Qy 1 MethisprothrLeuSerArgHisLeuProIleArgGlyLysProLysArgLeu 20
Db |||||
Qy 61 AAAATCCCTTAAAGCTTCGCCTCCATCGCGTGTTCCTTACTCTGTCAATAACCTCTCAG 120
Db |||||
Qy 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40
Db |||||
Qy 121 ACTAATGGTATGCGCATAGGACAGCGCTGAACCTCCATAAACCTTATCTCTCACCTGG 180
Db |||||
Qy 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
Db |||||
Qy 181 TTAATTAATGACTCGGCACAGGTATTAAATATCAACACTCAAGGGAGGCTCTTTA 240
Db |||||
Qy 61 LeuIleThrAspSerGlyThrGlyLeuAsnIleAsnSerThrGlnGlyAlaProLeu 80
Db |||||
Qy 241 GGAACCTGGTGGCCTGATCTATACGTTTGCCTCAGATCAGTTATTCTCTAGTCTG----- 294
Db |||||
```









## RESULT 14

Q8Q6Y7\_9GAMR  
 ID Q8Q6Y7\_9GAMR PRELIMINARY; PRT; 659 AA.  
 AC Q8Q6Y7;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Envelope glycoprotein.  
 GN Name=env;  
 OS Porcine endogenous retrovirus.  
 OC Viruses; Retrod viruses; Retroviridae; Gammaretrovirus;  
 OC 1-Mammalian type C virus group.  
 OX NCBI\_TaxID=61673;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=21851116; PubMed=11861871;  
 RX DOI=10.1128/JVI.76.6.3045-3048.2002;  
 RA Oldmixon B.A., Wood J.C., Ericsson T.A., Wilson C.A.,  
 RA White-Scharf M.E., Andersson G., Greenstein J.L., Schuurman H.J.,  
 RA Patience C.;  
 RT "Porcine endogenous retrovirus transmission characteristics of an  
 RT inbred herd of miniature swine.";  
 RL J. Virol. 76:3045-3048(2002).  
 DR EMBL; AF417231; AAL87626.1; -; Genomic\_DNA.  
 DR HSP; P03385; IMOF.  
 DR SMR; Q8Q6Y7; 509-561.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR002050; Env\_polyprotein.  
 DR InterPro; IPR008981; FmVlVrecept-bind.  
 DR Pfam; PF00429; Tlv\_coat; 1.  
 KW Envelope protein.  
 SQ SEQUENCE 659 AA; 73867 MW; 7B6A58D11541CA06 CRC64;

## Alignment Scores:

Pred. No.: 7,51e-234 Length: 659  
 Score: 2995.50 Matches: 569  
 Percent Similarity: 90.1% Conservative: 25  
 Best Local Similarity: 86.3% Mismatches: 44  
 Query Match: 86.4% Indels: 21  
 DB: 2 Gaps: 6

US-10-723-552-3\_copy\_5620\_7533 (1-1914) x Q8Q6Y7\_9GAMR (1-659)

QY 1 ATGCATCCACGTTAAACCGCGCCACTCCCGATTCCGGGTGGAAGCGAAAGACTG 60  
 DB 1 MethisProThrLeuSerArgArgHisLeuProLeuArgGlyGlyProLeuValArgLeu 20  
 QY 61 AAAATCCCTTAAGCTTCGCTCCATCGCGGTTCCTTACTCTGTCAATAACCTCTCAG 120  
 DB 21 LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
 QY 121 ACTAATGGTATCGCATAGGACAGCAGCTGACCTCCCATAAACCTTATCTCTCAGCTGG 180  
 DB 41 IleAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60  
 QY 181 TTAATTAAGTCCCGCACAGGTATTAAATATCAACACACTCAAGGGAGGCTCCCTTA 240  
 DB 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80  
 QY 241 GGAACCTGGTGGCTGATCTATAGTTTGCTTCAGATCAGTTATTCCTAGTCTG----- 294  
 DB 81 GlyThrTrpTrpProGluLeuLeuValCysLeuArgSerValIleProGlyLeuAsnAsp 100  
 QY 295 ----ACCTCACCCAGATATCTCCATGCTCAGGATTTTATGTTGCCAGGACACCA 351  
 DB 101 GlnAlaThrProProAspValLeuArgAlaTrpGlyPheTrpValCysProGlyProPro 120  
 QY 352 AATAATGGAACATTCGCGAATCCAGAGATTTCTTTGTAAACAATGGAAGTGTGA 411  
 DB 121 AsnAsnGluGluTrpCysGlyAsnProGlnAspPheCysLysGlnTrpSerCysVal 140

QY 412 ACCTCTAATGATGATATTGGAAATGGCAACCTCTCAGCAGGATAGGTAAAGTTTCT 471  
 DB 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160  
 QY 472 TATGCTCAACCTATACAGCTCTGGACAAATTAATTAC-----CTGACC 516  
 DB 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180  
 QY 517 TGG-----ATTGAAGACTGGAAGCCCAAGTGCCTCTCTCTCAGAC 555  
 DB 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200  
 QY 556 CTAGATTACCTAAATAAGTTTCACCTGAGAAAGGAAACAAAGAAAATATCTCTAAATGG 615  
 DB 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrp 220  
 QY 616 GTAAATGGTATGCTTGGGGAATGGTATATATGAGGCTCGGTAAACAACACAGGCTCC 675  
 DB 221 ValAsnGlyMetSerTrpGlyIleValTyrGlySerGlyArgLysGlySer 240  
 QY 676 ATTCTAATCTATCGCCTCAAAATA---AACCAGCTGGAGCCTCCAATGGCTATAGGACCA 732  
 DB 241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProProValAlaIleGlyPro 260  
 QY 733 AATACGGTCTTGACGGGTCAAAGACCCCAACCCAA-----GGACCAAGGACCA 780  
 DB 261 AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGluArgProSerProAsnPro 280  
 QY 781 TCCTCT---NACATAACTTCTGGATCAGACCCCACTAGTCTAACACAGCAGCTAAATG 837  
 DB 281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300  
 QY 838 GGGGCAAACTTTTAGCCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCACGACTCCA 897  
 DB 301 GlyAlaLysLeuPheAsnLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320  
 QY 898 GAGGCTACCTCTTTTGTGCTATGCTTAGCTTCGGGCCCACTTACTATGAAGGAATG 957  
 DB 321 GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMet 340  
 QY 958 GCTAGAAGAGGGAATTCATCTGACAAAAGAACATAGACACCACTACATCGAGTGGATCC 1017  
 DB 341 AlaArgGlyGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySer 360  
 QY 1018 CAAATAAGCTTACCTTACTAGGTTTCTGAAAAGCGACCTGCATAGAAAGGTCCCC 1077  
 DB 361 GluAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValPro 380  
 QY 1078 CCATCCCAACCAACCTTTGTAAACCACTGAAGCCTTTAATCAAACTCTGAGAGTCAA 1137  
 DB 381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGln 400  
 QY 1138 TATCTGTTACCTGGTTATGACAGGTGGTGGCATGTAATGTAATGTAATGTAATGTAATG 1197  
 DB 401 TyrLeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVal 420  
 QY 1198 TCCACCTTGGTTTAAACCAAACTTAAAGATTTTTCATTATGTCCTCAAAATGTTCCCCGA 1257  
 DB 421 SerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArg 440  
 QY 1258 GTGTATTACTATCCCGAAAAGCAATCTTGTATGAATATGACTACAGAAATCATCGACAA 1317  
 DB 441 ValTyrTyrTrpProGluLysAlaValLeuAspGluTyrAspTyrArgAsnHisArgGln 460  
 QY 1318 AAGAGAGACCCATATCTCTGACACTTCTGTGATGCTCGGACTTGGAGTGGGCGAGGT 1377  
 DB 461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGly 480  
 QY 1378 GTAGGAACAGGAACAGCTGCCCTGGTCACGGGACACAGCAGCTAGAAAACAGACTTAGT 1437  
 DB 481 ValGlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSer 500  
 QY 1438 AACCTACATCGAATTGTAAACAGAAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCTG 1497

```
Db 501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520
Qy 1498 GAGGAATCCCTAACTCTTATCTGAAGTAGCTCTACAGATAGAGAAGGTTAGATTTA 1557
Db 521 GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 540
Qy 1558 TTATTTCTAAAGAGGAGGATTATGCTAGCCTTGAAGGAGGAATGCTGTTTATGTTG 1617
Db 541 LeuPheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGlyCysPheTyrVal 560
Qy 1618 GATCATTCAGGGGCCATCAGAGACTCCATGAACAGCTTACAGAAAGGTTGGAGAGCGT 1677
Db 561 AspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArg 580
Qy 1678 CGAAGGGAAGAACTACTCAAGGGTGGTTTGAGGGATGTTCAACAGTCTCTTTGG 1737
Db 581 ArgArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrp 600
Qy 1738 TTGGCTACCTTCTTCTTAAACAGTAAATGGCTTACAGAGCCCTTAATAGTCCCTCTCTTACTACA 1797
Db 601 LeuAlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThr 620
Qy 1798 GTTGGCCATGTAATTATTAACAAGTTAATGGCTTCAATAGAGAACGAATAAGTGCAGTC 1857
Db 621 ValGlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaVal 640
Qy 1858 CAGATCATGCTAGTACAGACAGTACCAAGCCCGTCTACAGAGGAAGCTGGCCGC 1914
Db 641 GlnIleMetValLeuArgGlnGlnTrpGlnSerProSerArgGluAlaGlyArg 659

RESULT 15
Q6UQ77_9GAMR PRELIMINARY; PRT; 659 AA.
ID Q6UQ77_9GAMR AC Q6UQ77;
DC 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Envelope.
GN Name=env;
OS Porcine endogenous retrovirus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus;
OC 1-Mammalian type C virus group.
OX NCBI_TaxID=61673;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=14963152; DOI=10.1128/JVI.78.5.2502-2509.2004;
RA Scobie L., Taylor S., Wood J.C., Suling K.M., Quinn G., Meikle S.,
RA Patience C., Schuurman H.J., Onions D.E.;
RT "Absence of replication-competent human-tropic porcine endogenous
RT retroviruses in the germ line DNA of inbred miniature swine."
RL J. Virol. 78:2502-2509(2004).
DR EMBL; AY364236; AAR16717.1; -; Genomic_DNA.
DR HSSP; P03390; 1AOL.
DR SMR; Q6UQ77; 509-561.
DR GO; GO:0039028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env polyprotein.
DR Pfam; PF00429; TUV coat; 1.
SQ SEQUENCE 659 AA; 74068 MW; 466F54D8CF93014F CRC64;

Alignment Scores:
Pred. No.: 4,87e-230 Length: 659
Score: 2948.50 Matches: 562
Percent Similarity: 89.1% Conservative: 25
Best Local Similarity: 85.3% Mismatches: 51
Query Match: 85.0% Indels: 21
DB: 2 Gaps: 6

US-10-723-552-3_COPY_5620_7533 (1-1914) x Q6UQ77_9GAMR (1-659)
Qy 1 ATGCATCCACGCTTAACCGCGCCACCTCCCGATTCCGGGTGGAAAGCCGAAAGACTG 60
||||| :|||:|||||
```

```
Db 1 MetHisSerThrLeuSerArgArgHisPheProIleArgGlyGlyLysProLysArgLeu 20
Qy 61 AAAATCCCTTAAGCTTCGGCTCCATCGCTGGTTCCTTACTCTGTCAATAACCTCTCAG 120
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
Qy 121 ACTAATGCTATGCTAGGAGAGCGCTGAACCTCCCATAAACCTTATCTCTCACCTGG 180
Db 41 ValAsnGlyLysArgLeuValAspCysProAsnSerHisLysProLeuSerLeuThrTrp 60
Qy 181 TTAATTACTGATCCGCGCACAGGTATTAATATCAACAACACTCAAGGAGGAGCTCTTTA 240
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
Qy 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTCAGATCAGTATTATCTAGTCGTG----- 294
Db 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
Qy 295 ---ACCTCACCCCGAGATATCTCCATGCTCACCAGATTATGTTTGGCCAGGACCACCA 351
Db 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
Qy 352 AATAATGAAACAACTCGGAAATCCAGAGATTCTTTTCTTAAACAATGAACCTGTGTA 411
Db 121 AsnAsnGluGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
Qy 412 ACCTCTAATGATGATATTCGAAATGGCAACCTCTCAGCAGGATAGGTAAGTTTCT 471
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160
Qy 472 TATGTCAAACCTATACCAGCTCTGGCAATTTAATTAC-----CTGACC 516
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
Qy 517 TGG-----ATTAGAACTGGAAGCCCAAGTGTCTCTCTTCAGAC 555
Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
Qy 556 CTAGATTACCTAAATAAGTTTCACTCAGAAAGGAACAAAGAAATATCTCTAAATGG 615
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrp 220
Qy 616 GTAAATGCTATGCTTGGGAATGCTATATTATGAGGCTCGGTTAAACCAACAGGCTCC 675
Db 221 ValAsnGlyMetSerTrpGlyIleValTyrTyrGlySerGlyArgLysGlySer 240
Qy 676 ATTCCTAATATTCGCTCAAAATA---AACAGCTGGAGCTCCAAATGGCTATAGACCA 732
Db 241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProProValAlaIleGlyPro 260
Qy 733 AATAGGCTCTTGCGGGTCAAGAACCCCAACCCCAAGGA-----CCAGGACCA 780
Db 261 AsnLysGlyLeuAlaLysGlnGlyProProIleGlnGluArgProPheProAsnPro 280
Qy 781 TCCTCT---AACATAACTTCTGGATCAGACCCCACTAGTCTAACAGCACCACTTAAATG 837
Db 281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300
Qy 838 GGGGCAAACTTTTTCAGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCACGACTCCA 897
Db 301 GlyAlaLysLeuPheAsnLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320
Qy 898 GAGCTACCTCTTCTTGTGGTATGCTTAGCTTCGGGCCCACTTACTATCATGAGGAATG 957
Db 321 GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMet 340
Qy 958 GCTAGAGAGGGAATTCATATGTCAAAAGAAACATAGAGCAATGACATGCGGATGCC 1017
Db 341 AlaArgGlyGlyArgPheTyrValThrLysGluHisArgAspGlnCysAlaTrpGlySer 360
Qy 1018 CAAATAAGCTTACCTTACTAGGTTTCTGGAAAAGGCACCTGCATAGGAAAGGTTCC 1077
Db 361 GlnAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyMetValPro 380
```

QY 1078 CCATCCCAACCAACCTTTGTAACCACTGAAGCCTTTTAATCAAACTCTGAGAGTCAA 1137  
Db |||||||  
381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400  
|||:|||||  
QY 1138 TATCTGGTACCTGGTTATGACAGGTGGTGGCATGTAACTACTGGATTAAACCCCTTGTT 1197  
Db |||||||  
401 TyrLeuValProGlyTyrAspArgTrpAlaCysAsnThrGlyLeuThrProCysVal 420  
|||||  
QY 1198 TCCACCTTGGTTTTTAACCAAACTAAAGATTTTTGCATTATGGTCCAAATTTGCCCGA 1257  
Db |||||||  
421 SerThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArg 440  
|||||  
QY 1258 GTGTATTACTATCCGAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAA 1317  
Db |||||||  
441 ValTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGln 460  
|||||  
QY 1318 AAGAGAGAACCCATATCTCTGACACTGCTGCTGATGCTCGGACTTGGAGTGGCAGCAGGT 1377  
Db |||||||  
461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGly 480  
|||||  
QY 1378 GTAGGAACAGGAACAGCTGCCTGTGTACGGGACCACAGCAGCTAGAAACAGGACTTAGT 1437  
Db |||||||  
481 ValGlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSer 500  
|||||  
QY 1438 AACCTACATCGAATTGTTAACAGAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACCTG 1497  
Db |||||||  
501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520  
|||||  
QY 1498 GAGGAATCCCTAACCTCTTATCTGAGTAGTCTCTACAGATAGAGAGGTTAGATTTA 1557  
Db |||||||  
521 GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 540  
|||||  
QY 1558 TTATTTCTAAAGAGGAGGATTATGTAGCTTGAAGGAGGAATGCTGTTTTTATGTG 1617  
Db |||||||  
541 LeuPheLeuLysGluGlyLeuCysValAlaLeuLysGluGluCysPheTyrVal 560  
|||||  
QY 1618 GATCATTGAGGGCCATCAGAGACTCCATGAACAGCTTAGAGAAAGGTTGAGAAGCGT 1677  
Db |||||||  
561 AspHisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArg 580  
|||||  
QY 1678 CGAAGGGAAGGAAGAACTACTCAAGGGTGGTTTGAAGGATGGTTCAACAGGTCTCTTGG 1737  
Db |||||||  
581 ArgArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrp 600  
|||||  
QY 1738 TTGGCTACCTTACTTCTGCTTTAACAGGACCTTATAGTCTCTCTCTCTACTCACA 1797  
Db |||||||  
601 LeuAlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThr 620  
|||||  
QY 1798 GTTGGGCCATGTATTATTAAACAGTTAATTCCTTTCATTAGAGAACGAATAAGTGCAATC 1857  
Db |||||||  
621 ValGlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaVal 640  
|||||  
QY 1858 CAGATCATGTGTACTAGACAACAGTACCAGCCGCTCTAGCAGGGAAGCTGGCCGC 1914  
Db |||||||  
641 GlnIleMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 659  
|||||

Search completed: February 14, 2006, 14:51:01  
Job time : 448.173 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 12:52:14 ; Search time 50.9239 Seconds  
(without alignments)  
3302.856 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_5620\_7533

Perfect score: 3468

Sequence: 1 ATGCATCCACGTTAAACCG.....CTACGAGGAGAGCTGGCCGC 1914

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/abs/ABSSWEB\_spool/US10723552/runat\_14022006\_125139\_12833/app\_query.fasta\_1  
-DB-A Geneseq -QFT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=dits -START=1 -END=-1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs02p  
-USER=US10723552 @CGN 1.1.734 @runat\_14022006\_125139\_12833 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WAEI TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A Geneseq 21:\*  
1: geneseqp1980s:\*  
2: geneseqp1980s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3425	98.8	638	2 AAW32098	Miniature
2	3425	98.8	638	4 AAB70633	Porcine e
3	3425	98.8	638	8 AAB73287	Retroviral
4	3425	98.8	638	8 ADS73448	Swine ret
5	3420	98.6	638	4 AAB35114	PERV-C en
6	3407	98.2	638	9 ADY28028	Porcine e
7	3407	98.2	638	9 ADY28030	Porcine e
8	3407	98.2	638	9 ADY28036	Porcine e
9	3407	98.2	638	9 ADY28041	Porcine e

10	3407	98.2	638	9	ADY28038	Porcine e
11	3402	98.1	638	4	AAB70634	Porcine e
12	3402	98.1	666	8	ADS73449	Swine ret
13	3318	95.7	678	9	ADY28032	Porcine e
14	3318	95.7	678	9	ADY28034	Porcine e
15	3147.5	90.8	653	6	ABB82635	PERV enve
16	3135.5	90.4	653	6	ABB82636	PERV-1-15
17	2940.5	84.8	659	4	AAB35112	PERV-1-15
18	2879.5	83.0	661	9	ADY28039	Porcine e
19	2874.5	82.9	660	9	ADY28012	Porcine e
20	2871.5	82.8	660	9	ADY28004	Porcine e
21	2866.5	82.7	660	2	AAW85453	Pig endog
22	2866.5	82.7	660	4	AAW85453	Pig endog
23	2866.5	82.7	660	4	AAW85453	Pig endog
24	2851	82.2	678	9	ADY28032	Porcine e
25	2849.5	82.2	660	9	ADY28008	Porcine e
26	2849.5	82.2	660	9	ADY28014	Porcine e
27	2835	81.7	678	9	ADY28006	Porcine e
28	2755.5	79.5	657	2	AAW32095	Porcine e
29	2755.5	79.5	657	8	ADS73445	Swine ret
30	2755.5	79.5	1069	4	AAB73284	Defective
31	2465	71.1	678	9	ADY28022	Porcine e
32	2455	70.8	678	9	ADY28020	Porcine e
33	2432	70.1	657	9	ADY28026	Porcine e
34	2432	70.1	657	9	ADY28018	Porcine e
35	2432	70.1	658	9	ADY28040	Porcine e
36	2422	69.8	657	2	AAW85452	Pig endog
37	2422	69.8	657	4	AAB70635	Porcine e
38	2422	69.8	657	9	ADY28024	Porcine e
39	2422	69.8	657	9	ADY28016	Porcine e
40	2395	69.1	656	2	AAW39274	Porcine e
41	2353	67.8	656	2	AAW39273	Porcine e
42	2353	67.8	656	4	AAB70636	Porcine e
43	1989	57.4	540	6	ABP58412	Porcine e
44	1985	57.2	540	6	ABP58415	Porcine e
45	1985	57.2	540	6	ABP58413	Porcine e

#### ALIGNMENTS

RESULT 1

AAW32098

ID AAW32098 standard; protein; 638 AA.

XX

AC AAW32098;

XX

DT 27-AUG-2003 (revised)

DT 09-FEB-1998 (first entry)

XX

DE Miniature swine retrovirus ENV protein.

XX

KW Retrovirus; porcine; ENV protein; xenotransplantation; infectious; provirus; organ transplant; donor; activated virus; PCR.

XX

OS Pig endogenous retrovirus.

XX

PH Key Location/Qualifiers

FT Protein 1..638

FT /label= ENV\_protein

XX

PN WO921836-A1.

XX

PD 19-JUN-1997.

XX

PF 13-DEC-1996; 96WO-US019680.

XX

PR 14-DEC-1995; 95US-00572645.

XX

PA (GENO ) GEN HOSPITAL CORP.

XX

PI Fishman JA;

XX

DR WPI; 1997-332804/30.

DR N-PSDB; AAT74884.  
 XX New nucleic acid from porcine retroviruses - used for detecting viruses  
 PT in transplant or other tissue and for assessing risk of transmitting  
 PT infection to graft recipient.  
 XX  
 PS Claim 22; Fig 3; 128pp; English.

XX This is a porcine retrovirus from miniature swine containing the coding  
 CC region for a putative viral ENV protein. This sequence and PCR fragments  
 CC generated from the sequence (see AAT74812-T74882) could be used to screen  
 CC organs for porcine retroviruses prior to xenotransplantation.  
 CC Transplantation can increase the likelihood of retroviral activation if  
 CC intact and infectious proviruses are present. The porcine retroviral  
 CC sequence can be used to generate probes to determine the level (e.g. copy  
 CC number) of intact (i.e. potentially replicating) porcine provirus  
 CC sequences in a strain of xenograft transplantation donors. It can be used  
 CC to detect mutations, genetic lesions or viral recombinants and also to  
 CC determine the histological localisation of activated retrovirus. Using  
 CC Polymerase Chain Reaction DNA Quantitation (PQD) on blood mononuclear  
 CC cells, infectivity titration and susceptibility testing can be performed.  
 CC ultimately animal donors without intact porcine retroviral sequences or a  
 CC lower copy number of viral elements could be selected. (Updated on 27-AUG  
 CC -2003 to correct OS field.)  
 XX

SQ Sequence 638 AA;

#### Alignment Scores:

Pred. No.: 2.8e-291 Length: 638  
 Score: 3425.00 Matches: 638  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 98.8% Indels: 0  
 DB: 2 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x AAW32098 (1-638)

QY 1 ATGCATCCAGCTTAACCGCGCCACCTCCGATTCGGGTGGAAAGCGAAGACTG 60  
 DB 1 MethisProThrLeuAsnArgHisLeuProIleArgGlyLysProLysArgLeu 20  
 QY 61 AAAATCCCTTAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGCTCAATAACCTCTCAG 120  
 DB 21 LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
 QY 121 ACTAATGGTATCGCATAGGACAGCTGCACTCCCATAAACCTTATCTCTCACCTGG 180  
 DB 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
 QY 181 TTAATTACTGACTCCGGCAGAGTATTAATATCAACACACTCAAGGGAGGCTCCTTTA 240  
 DB 61 LeuileThrAspSerGlyThrGlyIleAsnileAsnAsnThrGlnGlyAlaProLeu 80  
 QY 241 GGAACCTGGTGGCTGATCTACGTTTGCCTCCAGATCAGTATTCTTCTGACCTCA 300  
 DB 81 GlyThrTrpProAspLeuThrValCysLeuArgSerValIleProSerLeuThrSer 100  
 QY 301 CCCCCAGATATCTTCATGCTCACGGATTTATGTTGCCAGGACACCCAAATAATGGA 360  
 DB 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120  
 QY 361 AAACNTCCGGAAATCCAGAGATTTCTTTGTAAACATGGAAGTGTAACTCTTAAT 420  
 DB 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140  
 QY 421 GATGATATGGAATGCGCAACCTCTCAGCAGGATAGGTAAGTTTCTTATGCTCAAC 480  
 DB 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
 QY 481 ACCTATACCAGCTCTGGCAATTTAATTAACCTGACCTCGATTAGAACCTGGAGCCCCAAG 540  
 DB 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180

QY 541 TGCTCTCTCTTCAGACCTAGATTACCTAAAAATAAGTTTCTACTGAGAAAGGAAAAACAAGAA 600  
 DB 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
 QY 601 AATATCTCTAAATGGGTAAATGGTATCTCTGGGGAATGGTATATATATGAGGCTCGGT 660  
 DB 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGly 220  
 QY 661 AAACACACAGCTCCATCTTAACTATTCGCTCAAAATAAACACAGCTGGAGCTCCCAATG 720  
 DB 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
 QY 721 GCTATAGACCAATAATACGGTCTTCGACGGGTCAAGAGACCCCAACCCCAAGACACAGGACCA 780  
 DB 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
 QY 781 TCCTCTAACATACTCTGGATCAGACCCCACTGAGTCTAACACAGCAGCTAAATAATGGG 840  
 DB 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
 QY 841 GCAAACTTTTACCTCATCCAGGAGCTTTCAAGCTCTTAAGTCTTAAGTCTCCAGCTCCAGAG 900  
 DB 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGln 300  
 QY 901 GCTACCTCTCTCTGCTGCTATGCTTAGCTTCGGGCCCCACCTTACTATGAAGGAATGGCT 960  
 DB 301 AlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMetAla 320  
 QY 961 AGAAGAGGGAATTCATATGTGACAAAAACATAGACCAATGCATGCATGGGGATCCCAA 1020  
 DB 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
 QY 1021 AATAGCTTACCTTACTGAGTTCCTGGAAAGGACCTGTCATAGGAAGAGTTCCCCCA 1080  
 DB 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
 QY 1081 TCCCAACCAACCTTTGTAAACCACTCAAGACCTTTAATCAAACTCTCGAGAGCAATAT 1140  
 DB 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
 QY 1141 CTGCTACTCTGTTATGACAGGTGTGGGCATGTAACTACTGGATTAAACCCCTTGTGTTCC 1200  
 DB 381 LeuValProGlyTyrAspArgTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
 QY 1201 ACCTTGGTTTAAACCAACTAAAGATTTTGCATTTGCTCCAAATTTGTTCCCGAGTG 1260  
 DB 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
 QY 1261 TATTACTATCCGAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAG 1320  
 DB 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
 QY 1321 AGAGAACCCTATCTCTGACACTTGTGTGATCTCGGACTTGGAGTGGCAGCGGTGTA 1380  
 DB 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
 QY 1381 GGAACAGAACAGCTGCCCTGGTCACGGGACCAAGACGCTAGAAACAGGCTTAGTAAC 1440  
 DB 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
 QY 1441 CTACATCGAATTTGTAACAGAGATCTCCAAGCCCTAGAAAAATCTGTCTAGTAACCTCGAG 1500  
 DB 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
 QY 1501 GAATCCCTTAACCTCTTATCTGAAGTAGTCTCTCAGAAATAGAGAGGTTAGATTATTA 1560  
 DB 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
 QY 1561 TTTCTAAAGAGGAGGATTTATGTAGCTTGGAGGAGGAATCTGTGTTTATGTGTGAT 1620  
 DB 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAsp 540  
 QY 1621 CATTGAGGGGCCATCAGAGACTCCATGAACCAAGCTTTAGAGAAAGGTTGGAGAAAGCGTGA 1680





```
QY 961 AGAAGAGGAAATTCATGTGACAAAAGAACATAGAGACCAATGCACATGGGATCCCAA 1020
Db |||||||
QY 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340
Db |||||||
QY 1021 AATAAGCTTACCTTACTAGGTTTTCGAAAAGGCACCTGCATAGGAAAGTTTCCCCCA 1080
Db |||||||
QY 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360
Db |||||||
QY 1081 TCCACCAACACCTTTGTAACACACTGAAGCTTTAATCAACCTCTGAGAGTCAATAT 1140
Db |||||||
QY 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380
Db |||||||
QY 1141 CTGGTACCTGTTATGACAGGTGGGCGCATGTATATCTGCAATTAACCCCTTGTTGTTCC 1200
Db |||||||
QY 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400
Db |||||||
QY 1201 ACCTTGGTTTTAAACCAAACTAAAGATTTTGCATTATGGTCCAAATTTGTTCCCGAGTG 1260
Db |||||||
QY 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420
Db |||||||
QY 1261 TATTACTATCCGAAAAGCAATCCTTCATGAATATGACTACAGAAATCATCGACAAG 1320
Db |||||||
QY 421 TyrTyrTrpProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440
Db |||||||
QY 1321 AGAGAACCATATCTCTCACACTTCTGTCGATGCTCGACTTGGAGTGGCAGCAGGTGTA 1380
Db |||||||
QY 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460
Db |||||||
QY 1381 GGAACAGGAACAGCTGCCCTGGTTCACGGGACCACAGCAGCTAGAACAGGACTTAGTAAC 1440
Db |||||||
QY 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
Db |||||||
QY 1441 CTACATCAATTTAAACAGAGATCTCCAGCCCTAGAAAATCTGTCAGTAACTCGAG 1500
Db |||||||
QY 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500
Db |||||||
QY 1501 GAATCCCTAACTCTTATCTGAACTAGTCTTACAGAAATAGAGAGGTTAGATTTATTA 1560
Db |||||||
QY 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeu 520
Db |||||||
QY 1561 TTTCTAAAGAGGAGGATTTATGTGTAGCTTTGAAGGAGGAATGCTGTTTTATGTGGAT 1620
Db |||||||
QY 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAsp 540
Db |||||||
QY 1621 CATTCAGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAGGTTGGAGAGGTCGA 1680
Db |||||||
QY 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560
Db |||||||
QY 1681 AGGGAAGGAAACTPACTCAAGGGTGGTTTGAGGGATGGTTCAACAGGTCTCTTTGGTTG 1740
Db |||||||
QY 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeu 580
Db |||||||
QY 1741 GCTACCTTACTTTCTGCTTTAACAGGACCCTTAATAGTCCTCTCTCTTACTCACAGTT 1800
Db |||||||
QY 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600
Db |||||||
QY 1801 GGGCCATGTTATTAAACAGTTAATTCGCTTCATTAGAGAACGAAATAGTCAGTCCAG 1860
Db |||||||
QY 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620
Db |||||||
QY 1861 ATCATGGTACTTAGACACAGTACCAGAGCCGCTCTAGCAGGAGCTGGCGCG 1914
Db |||||||
QY 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg 638
Db |||||||
```

## RESULT 3

AAB73287

ID AAB73287 standard; protein; 638 AA.

AC AAB73287;

XX

DT 23-MAY-2001 (first entry)

XX

DE Retroviral protein #3 found in miniature swine.

```
XX Retrovirus; graft transplantation; xenotransplantation; miniature swine.
XX Unidentified.
XX US6190861-B1.
XX 20-FEB-2001.
XX 13-DEC-1996; 96US-00766528.
XX 14-DEC-1995; 95US-00572645.
XX (GEHO ) GEN HOSPITAL CORP.
XX Fishman JA;
PI WPI; 2001-256211/26.
DR N-PSDB; AAF77727.
XX Assessing risk of endogenous retroviruses in clinical practice and in
PT xenotransplantation, comprises using probe sequences derived from swine
PT or miniature swine retroviral genome.
XX Disclosure; Fig 3; 127pp; English.
XX The present invention relates to a method for screening a cell or tissue
CC for the presence or expression of a retrovirus (RV), comprising
CC contacting a target nucleic acid from the cell or tissue with a second
CC nucleic acid from the present invention (e.g. AAF77727 or a fragment
CC thereof). The method is useful for RV detection and to assess graft
CC transplantation risk. Screening of animals allows the elimination of
CC donors with active replication of known viruses. Inactive proviruses can
CC be detected and inactivated, allowing identification and elimination of
CC potential human pathogens derived from swine in a manner not possible in
CC the outbred human organ donor population and is important to the
CC development of human xenotransplantation
XX Sequence 638 AA;
```

## Alignment Scores:

Pred. No.:	2,86-291	Length:	638
Score:	3425,00	Matches:	638
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	98.8%	Indels:	0
DB:	4	Gaps:	0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x AAB73287 (1-638)

QY	1	ATGCATCCACGTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAAGCCGAAAGACTG	60
Db	1	MethHisProThrLeuAsnArgHisLeuProIleArgGlyGlyLysProLysArgLeu	20
QY	61	AAAAATCCCTTAAAGCTTCGCCTCCATCGCGTGGTTCCTTACTCTGTCAATAACCTCTCAG	120
Db	21	LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln	40
QY	121	ACTAATGATGTCGATAGGACAGCAGCTGAACCTCCCAATAACCTTATCTCTCACCTGG	180
Db	41	ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp	60
QY	181	TTAATTACTGACTCGGCACAGGTATTATATCAACACTCAAGGGAGGCTCTTTTA	240
Db	61	LeuIleThrAspSerGlyThrGlyIleAsnIleAsnThrGlnGlyGluAlaProLeu	80
QY	241	GGAACTGGTGGCTGTATAGCTTTCCTCAGATCAGTATTTCCTAGTCTGACCTCA	300
Db	81	GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer	100
QY	301	CCCCCAGATATCTCCATGTCTCAGGATTTTATGTTTGGCCAGGACCAACCAATAATGGA	360
Db	101	ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly	120

QY 361 AAACATTCGGGAATCCAGAGATTCTTTTGTAAACAAATGGAACCTGTGTAACCTCTAAT 420  
Db |||||  
121 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 140  
QY 421 GATGGATATCGGAATGCGCAACCTCTCAGCAGGATAGGGTAAGTTTCTTATGTCAAC 480  
Db |||||  
141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
QY 481 ACCTATACAGCTCTGGACAATTTAATTAACCTCGACTGATTTAGAACTGGAGGCCCAAG 540  
Db |||||  
161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180  
QY 541 TGCTCTCCTTCAGACCTAGATTACCTAAATAAAGTTTCACTGAGAAAGGAAACAAGAA 600  
Db |||||  
181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
QY 601 AATATCCTAAATGGTAAATGCTATGCTTGGGGAATGGTATATTATGAGAGGCTCGGT 660  
Db |||||  
201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGly 220  
QY 661 AAACAACAGAGCTCAATCTAATCTAATTCGCCTCAAAATAAACACAGCTGGAGCTCCAATG 720  
Db |||||  
221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
QY 721 GCTATAGGACCAATACGGTCTTGACGGGTCAAGACCCCAAGCCCAAGGACCAAGCA 780  
Db |||||  
241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
QY 781 TCCTCTAACAATCTCTGGATCAGACCCCACTGAGTCTAAGCAGCAGCACTAAATGGGG 840  
Db |||||  
261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
QY 841 GCAAACTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCACGACTCCAGAG 900  
Db |||||  
281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 901 GCTACCTCTCTCTGCTATGCTTAGCTTCGGGCCCCACCTTACTATGAGGAAATGGCT 960  
Db |||||  
301 AlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMetAla 320  
QY 961 AGAAGAGGAAATCAATGTGACAAAAGAACATGAGACCAATGCATGCGGATCCCAA 1020  
Db |||||  
321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 1021 AATAAGCTTACCTTACTGAGTTCGTGGAAAGGCACCTGCATAGAAAGGTTCCCCCA 1080  
Db |||||  
341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 1081 TCCCAACCAACCTTTGTAACCACTGAAGCCTTTAATCAAACTCTGAGAGTCAATAT 1140  
Db |||||  
361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
QY 1141 CTGTACTGTTATGACAGTGTGGCATGTAACTGATTAACCCCTTGTTGTTCC 1200  
Db |||||  
381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 1201 ACCTTGGTTTTTAACCAAACTAAAGATTTTGTCAATTATGTTCCAAATGTTCCCGAGTG 1260  
Db |||||  
401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 1261 TATTACTATCCGAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCCGACAAAAG 1320  
Db |||||  
421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
QY 1321 AGAAGACCCATATCTCTGACACTTGTGATGTCGGACTTGGAGTGGCAGCAGGTGTA 1380  
Db |||||  
441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
QY 1381 GGAACAGACAGCTGCCCTGCTCAGCGGACCAAGCAGCTAGAAACAGGACTTAGTAAC 1440  
Db |||||  
461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480

QY 1441 CTACATCGAATGTGAACAGAGAGATCTCCAAGCCCTAGAAAAATCTGTCAAGTAACCTGGAG 1500  
Db |||||  
481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 1501 GAATCCCTTAACCTCTTATCTGAAGTAGTCTCAGAGATAGAGAGGGTTAGATTATTA 1560  
Db |||||  
501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
QY 1561 TTTCTAAAAGAGGAGGATTATGTGTAGCCTTTGAAGGAGGAATGCTGTTTTTATCTGGAT 1620  
Db |||||  
521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAsp 540  
QY 1621 CATTACGGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGGTTGAGAGAGCGTCCA 1680  
Db |||||  
541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560  
QY 1681 AGGAAAAGGAACTACTCAAGGGTGTTCAGGGATGGTTCAACAGGTCTCTTTGGTTG 1740  
Db |||||  
561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeu 580  
QY 1741 GCTACCTACTTTCTGCTTTAAACAGGACCTTAATAGTCTCTCTCTCTTACTCACAGTT 1800  
Db |||||  
581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrVal 600  
QY 1801 GGGCCATGTATTATTAAACAAGTTAATTCCTTTCATTAGAGAACGAATAAGTCAGTCCAG 1860  
Db |||||  
601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgGlyIleSerAlaValGln 620  
QY 1861 ATCATGTACTTAGACACACAGTACCAAGGCCCTCTAGCAGGGAAGCTGGCCGC 1914  
Db |||||  
621 IleMetValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg 638

RESULT 4  
ADS73448  
ID ADS73448 standard; protein; 638 AA.  
XX ADS73448;  
XX  
XX  
DT 16-DEC-2004 (first entry)  
XX  
DE Swine retroviral env protein.  
XX  
KW Swine retroviral protein; immunosuppressive; gene therapy; env protein.  
XX  
XX Porcine endogenous retrovirus.  
OS  
XX US2004185435-A1.  
XX  
XX 23-SEP-2004.  
XX  
XX 26-NOV-2003; 2003US-00723552.  
XX  
XX 14-DEC-1995; 95US-00572645.  
XX 13-DEC-1996; 96US-00766528.  
XX 14-SEP-2000; 2000US-00661858.  
XX  
XX (GSHO ) GEN HOSPITAL CORP.  
XX  
XX Fishman JA;  
XX  
XX WPI; 2004-689179/67.  
XX N-PSDB; ADS73369.  
XX  
XX New porcine retroviral polypeptide encoded by a nucleic acid, useful in  
PT evaluating an immunosuppressive treatment for the ability to activate a  
PT retrovirus, such as an endogenous porcine retrovirus.  
XX  
XX PS Claim 8; Fig 3; 83pp; English.  
XX  
XX The present invention relates to the swine retroviral polypeptides and  
CC their encoding nucleic acids. The methods and compositions of the present  
CC invention are useful for screening a cell or tissue, e.g. a heart, lung,  
CC liver, bone marrow, kidney, brain cells, neural tissue, pancreas and

CC intestinal tissue xenograft, for the presence or expression of a swine or  
 CC miniature swine retrovirus or retroviral sequence. The invention is also  
 CC useful in evaluating an immunosuppressive treatment for the ability to  
 CC activate a retrovirus such as an endogenous porcine retrovirus. The  
 CC invention is also useful in gene therapy. The present sequence is the  
 CC swine retroviral env protein.

xx Sequence 638 AA;

SQ

Alignment Scores:

Pred. No.: 2.8e-291 Length: 638  
 Score: 3425.00 Matches: 638  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 98.8% Indels: 0  
 DB: 8 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x ADS73448 (1-638)

Qy 1 ATGATCCACGTTAAACCGCGGCACCTCCGATTCGGGTGGAAAGCCGAAAGACTG 60  
 Db 1 MethisProthrLeuAsnArgHisLeuProIleArgGlyGlyLeuProLysArgLeu 20  
 Qy 61 AAAATCCCTTAAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCAATAACCTCTCAG 120  
 Db 21 LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
 Qy 121 ACTAATGGTATGCGCATAGGAGACGCTGAACCTCCCATAAACCTTATCTCTCACCTGG 180  
 Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
 Qy 181 TTAATTACTGACTCCGGCACAGGTATTAATATCAACACTCAAGGGAGGCTCCCTTA 240  
 Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnThrGlnGlyGluAlaProLeu 80  
 Qy 241 GGAACCTCGTGGCTGATCTATAGCTTTCCTCCAGATCAGTATATTCCTAGTCTGACCTCA 300  
 Db 81 GlyThrTrpTrpProAspLeuTyValCysLeuArgSerValIleProSerLeuThrSer 100  
 Qy 301 CCCCAGATATCTCCATGTCACGGATTTATGTTTCCAGACACCAACCAATAATGGA 360  
 Db 101 ProProAspIleLeuHisAlaHisGlyPheTyValCysProGlyProProAsnAsnGly 120  
 Qy 361 AAACATTCGGAAATCCAGAGTTCTTTTGTAAACATGGAAGTGTAACTCTTAAT 420  
 Db 121 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 140  
 Qy 421 GATGATATTGGAATGGCAACCTCTCAGCAGGATAGGTAAGTTTCTTATGTCAAC 480  
 Db 141 AspGlyTyTrpLysTrpProThrSerGlnAspArgValSerPheSerTyValAsn 160  
 Qy 481 ACCTATACCAGCTCTGGACAATTTAATTTACCTGACCTCGATAGAACTGGAGCCCCAAG 540  
 Db 161 ThrTyThrSerSerGlyGlnPheAsnTyLeuThrTrpIleArgThrGlySerProLys 180  
 Qy 541 TGCTCTCTTCAGACCTAGATTACCTAAATAGTTTCACTGCAAGAGGAAACAGAA 600  
 Db 181 CysSerProSerAspLeuAspTyLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
 Qy 601 AATATCTTAAATGGGTAAGTGTATGCTTGGGAATGGTATATATGAGAGGCTCGGGT 660  
 Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyTrpGlyLysSerGly 220  
 Qy 661 AAACACAGGCTCCATTTCTAATTTACCTCGCTCAAAATAAACAGCTGGAGCTCCCAATG 720  
 Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
 Qy 721 GCTATAGACCAATACCGTCTTGACGGTCAAGACCCCAACCAAGGACCCAGGACCA 780  
 Db 241 AlaIleGlyProAsnThrValIleuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
 Qy 781 TCCTCTAACATAACTTCTGGATTCAGACCCCACTAGTCTTAACAGCACGACTAAATATGGG 840  
 Db

Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
 Qy 841 GCRAAACTTTTATGACCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCACAGCTCCAGAG 900  
 Db 281 AlalysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAenSerThrThrProGlu 300  
 Qy 901 GCTACCTCTTCTTGTGGCTATGCTTAGCTTCGGGCCACCTTACTATGAAGGAATGGCT 960  
 Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyTrpTyTrpGluGlyMetAla 320  
 Qy 961 AGAAGAGGAAATCAATGTGACAAAAGAACATAGACCAATGCACATGAGGATCCCAA 1020  
 Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
 Qy 1021 AATAAGCTTACCTTACTAGGTTTCTGGAAAGCGCACCTGCATAGGAAGGTTCCCCCA 1080  
 Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
 Qy 1081 TCCCAACCAACACCTTTCTTAACCACTGAAGCCCTTTAATCAAACTCTGAGAGTCAATAT 1140  
 Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTy 380  
 Qy 1141 CTGTACTCTGTTATGACAGGTGCTGGGCATGTAACTACTGGATTAAACCCCTTGTGTTC 1200  
 Db 381 LeuValProGlyTyTrpAspArgTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
 Qy 1201 ACCTTGGTTTTTAACCAACTAAAGATTTTGTGATTATGTCCTCAAACTGTTCCCGAGTG 1260  
 Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
 Qy 1261 TATTACTATCCGAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGCAAAAG 1320  
 Db 421 TyTrpTyTrpProGluLysAlaIleLeuAspGluTyTrpAspTyTrpAsnHisArgGlnLys 440  
 Qy 1321 AGAAGAACCATATCTCTGACACTGCTGATGTCGGAGTTCGGAGTTCGGAGTTCGGAGT 1380  
 Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
 Qy 1381 GGAACAGGACAGCTGCCCTGTCACGGGACCAAGAGCTAGAGAAAGAGGACTTAGTAAC 1440  
 Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
 Qy 1441 CTACATCGAATGTAAACAGAGATCTCAAGCCCTAGAAAAATCTGTCTAGTAACCTCGAG 1500  
 Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
 Qy 1501 GAATCCCTTAACCTCTTATCTGAAGTAGTCTCTAAGATAGAGAGGTTAGATTTATA 1560  
 Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
 Qy 1561 TTTCTAAAGAGGAGGATTTATGTCCTTGAAGGAGGATGCTGTTTATGTCGAT 1620  
 Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyValAsp 540  
 Qy 1621 CATTCAGGGGCATCAGAGACTCCATGAACAGCTTAGAGAAAGGTTTCGAGAAAGCGTCCA 1680  
 Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560  
 Qy 1681 AGGAAAGGAAACTACTCAAGGGTGGTTTTCAGGGATGGTTCAACAGGCTCTTTGGTTG 1740  
 Db 561 ArgGluLysGluThrThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeu 580  
 Qy 1741 GCTACCTCTCTTCTGCTTAAAGGACCTTAACTAGTCTCCCTCTGTTACTACAGTT 1800  
 Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
 Qy 1801 GGGCATCTATTATTAAACAGTTAATTCCTTCAATAGAGAACCAATTAAGTCAGTCCAG 1860  
 Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
 Qy 1861 ATCATGTGTACTAGAACACAGTACAAAGCCCTCTAGCAGGAGAGCTGGCCGC 1914  
 Db 621 IleMetValLeuArgGlnGlnTyTrpGlnSerProSerArgGluAlaGlyArg 638

## RESULT 5

AAB35114  
ID AAB35114 standard; protein; 638 AA.

XX AC AAB35114;

XX 27-MAR-2001 (first entry)

XX PERV-C env protein SEQ ID NO: 5.

XX Xenotransplantation; infectious agent; vaccine.

XX Porcine endogenous retrovirus.

XX WO200071726-A1.

XX 30-NOV-2000.

XX 24-MAY-2000; 2000WO-US014296.

XX 24-MAY-1999; 99US-0135631P.

XX (MAYO-) MAYO MEDICAL VENTURES.

XX Federspiel MJ;

XX WPI; 2001-032041/04.

XX Inhibiting or preventing infectious agent transmission in mammalian  
transplant recipients, by introducing recombinant DNA comprising DNA  
encoding extracellular proteins of the agent into donor cells, such as  
swine cells.

XX Disclosure; Page 95-97; 144pp; English.

XX The present invention provides a method to prevent the transmission of  
infectious agents during xenotransplantation. This involves introducing  
CC to donor swine cells a recombinant DNA encoding a peptide fragment from  
the infectious agent, and then introducing these cells into the  
transplant recipient

XX Sequence 638 AA;

## Alignment Scores:

Pred. No.: 7,69e-291 Length: 638  
Score: 3420.00 Matches: 637  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.8% Mismatches: 0  
Query Match: 98.6% Indels: 0  
DB: 4 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x AAB35114 (1-638)

Qy	1	ATGCATCCACGTTAAACCGCGCCACCTCCCGATTCGGGTGGAAAGCGAAAGACTG	60
Db	1	MethisProthrLeuAsnArgAgHisLeuProilleArgGlyGlyLysProLysArgLeu	20
Qy	61	AAATCCCTTAAGCTTCGGCTCCATCGCGTGGTTCCTACTCTGTCAATAACCTCTCAG	120
Db	21	LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln	40
Qy	121	ACTAATCGGTATGCGCATAGGAGACAGCTCAACTCCCATAAACCTTATCTCTCACTGG	180
Db	41	ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp	60
Qy	181	TTAATTACTGACTCCGCGCACAGGTATTATCAACACACACTCAAGGGGAGGCTCCCTTTA	240
Db	61	LeuileThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu	80
Qy	241	GGAACTGGTGGCTGATCTATAGCTTTGCTCAGATCAGTTATTCCTAGCTGACCTCA	300
Db	81	GlyThrTrpTrpProaspLeuThrValCysLeuArgSerValIleProSerLeuThrSer	100

Qy	301	CCCCAGATATCTCTCCATGCTCAGGATTTTATGTTTCCAGGACCAACCAATAATGGA	360
Db	101	ProProaspIleLeuHisAlaHisGlyPheTyValCysProGlyProProAsnAsnGly	120
Qy	361	AAACATTGCGGAATCCAGAGATTTCTTTTGTAAACAATGAACTGTGTAACTCTAAT	420
Db	121	LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn	140
Qy	421	GATGGATATTGGAATGCCAACCCTCTCAGCAGATAGGGTAAGTTTCTTATGTCAAC	480
Db	141	AspGlyTyTrpLysTrpProThrSerGlnAspArgValSerPheSerTyValAsn	160
Qy	481	ACCTATACCAAGCTCTGCACAAATTTAATTACTGACCTGGATTAGAACTGGAGGCCCAAG	540
Db	161	ThrTyThrSerSerGlyGlnPheAsnTyLeuThrTrpIleArgThrGlySerProLys	180
Qy	541	TGCTCTCTTCAGACCTAGATTACCTAAATAAGTTTCACTGAGAAGGAAAAACAAGAA	600
Db	181	CysSerProSerAspLeuAspTyLeuLysIleSerPheThrGluLysGlyLysGlnGlu	200
Qy	601	AATATCCTAAATGGGTAAATGGTATGCTCTTGGGAATGGTATATTATGAGGCTCGGT	660
Db	201	AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyTyGlyGlySerGly	220
Qy	661	AAACAACCAAGCTCCATTCTAATCTATTCGCTCAAAATAAACCAAGCTGGAGCCTCCAATG	720
Db	221	LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProwet	240
Qy	721	GCTATAGGACCAATACGGTCTTTGACGGGTCAAGACCCCAACCCCAAGGACCAAGACCA	780
Db	241	AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro	260
Qy	781	TCCTCTAACATACTCTTGATCAGACCCCACTGAGTCTAACACGACGACTAAATGGG	840
Db	261	SerSerAsnIleThrSerGlySerAspProThrGluSerSerThrThrLysMetGly	280
Qy	841	GCAAACTTTTACCTCATCCAGGAGCTTTTCAAGCTCTTAAGCTCCAGACTCCAGAG	900
Db	281	AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu	300
Qy	901	GCTACCTCTTCTTGGCTATGCTTAGCTTCGGGCCCACTTACTATGAAGGAATCGCT	960
Db	301	AlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyTyGlyGlyMetAla	320
Qy	961	AGAAGAGGGAATTCATGTGACAAAAGACATAGAGACCAATGCACATGGGATCCCAA	1020
Db	321	ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln	340
Qy	1021	AATAAGCTTACCTTACTGAGGTTCCTGGAAAAAGGCACCTGCATAGGAAAGTTCCCCCA	1080
Db	341	AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro	360
Qy	1081	TCCCAACCAACCTTTGTAAACCACTGAAGCCCTTTAATCAAACTCTGAGAGTCAATAT	1140
Db	361	SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTy	380
Qy	1141	CTGGTACCTGGTTATCAGAGGTGGCGATGTAATCTGGATTAAACCTTGTGTTC	1200
Db	381	LeuValProGlyTyAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer	400
Qy	1201	ACCTCTGTTTAAACCAAACTAAGATTTTTCATTATGTCCTCAAAATGTTTCCCGAGTG	1260
Db	401	ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal	420
Qy	1261	TATTACTATCCCGAAAAAGCAATTCCTTGATGATATGACTACAGAAATCATCGCAAAAG	1320
Db	421	TyrTyTyProGlyLysAlaIleLeuAspGluTyAspTyArgAsnHisArgGlnLys	440
Qy	1321	AGAAACCATATCTCTGACACTTCTGTGATGCTCGGACTTGGAGTGGCAGGAGGTGTA	1380
Db	441	ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal	460

```
QY 1381 GGAACAGGACAGCTGCTGCTACAGCGGACCACACAGCAGCTAGAAACAGGACTTAGTAAC 1440
Db |||||||
QY 461 GlyThrGlyThrAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
Db |||||||
QY 1441 CTACATCGAATTGTAACAGAAGACTCTCCAAGCCCTAGAAAAATCTGTCAGTAACCTGGAG 1500
Db |||||||
QY 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500
Db |||||||
QY 1501 GAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGATAGAGAAGGGTTAGATTATTA 1560
Db |||||||
QY 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520
Db |||||||
QY 1561 TTTCTAAAGAGGAGGAGTATGTGTAGCTTTGAAGGAGGAATGCTGTTTTTATGTGCGAT 1620
Db |||||||
QY 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysPheTyrValAsp 540
Db |||||||
QY 1621 CATTACAGGGCCATCAGAGACTCCATGAACAGCTTAGAGAAGGTTGGAGAAGCGTCCA 1680
Db |||||||
QY 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560
Db |||||||
QY 1681 AGGNAAGAAACTACTCAAGGTGGTTTGAGGATGGTTCACAGCTCTCTTGGTTG 1740
Db |||||||
QY 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeu 580
Db |||||||
QY 1741 GCTACCTACTTCTGCTTTAACAGAGCCCTTAATAGTCTCTCTCTCTACTCACAGTT 1800
Db |||||||
QY 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600
Db |||||||
QY 1801 GGGCATGTATTATTAAACAAGTTAATTCGCTTCATTAGAGAACGAATAAGTCAGTCCAG 1860
Db |||||||
QY 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620
Db |||||||
QY 1861 ATCATGGTACTTAGACAACAGTACCAAGCCCGTCTAGCAGGAAGCTGGCGGC 1914
Db |||||||
QY 621 IleMetValLeuArgGlnGlnTrpGlnSerProSerSerArgGluAlaGlyArg 638
Db |||||||
```

## RESULT 6

```
ADY28028
ID ADY28028 standard; protein; 638 AA.
XX
AC ADY28028;
XX
DT 19-MAY-2005 (first entry)
XX
DE Porcine endogenous retrovirus subtype C ancestral env protein, SEQ ID 32.
XX
KW Xenotransplantation; vaccine; acquired immune deficiency syndrome;
KW anti-hiv; immune disorder; HIV-1 infection; infection; PERV infection;
KW diagnosis; graft versus host disease; immunosuppressive;
KW envelope protein.
XX
OS Porcine endogenous retrovirus; subtype C.
OS Synthetic.
XX
FN WO2005019411-A2.
XX
PD 03-MAR-2005.
XX
PF 19-MAY-2004; 2004WO-US015709.
XX
PR 19-MAY-2003; 2003US-00441949.
XX
PA (AUCK-) AUCKLAND UNISERVICES LTD.
PA (UNIW ) UNIV WASHINGTON.
XX
PI Mullins JI, Rodrigo A, Ross HA;
XX
DR WPI; 2005-202625/21.
DR N-PSDB; ADY28027.
XX
PT New isolated ancestral viral nucleic acid sequence that is a determined
PT founder sequence of a highly diverse viral strain, subtype or group of an
PT endogenous retrovirus, useful in the prophylaxis of viral infection.
```

```
XX Claim 10; SEQ ID NO 32; 278pp; English.
PS
XX The invention relates to an isolated ancestral viral nucleic acid
CC sequence and its fragment, where the sequence is a determined founder
CC sequence of a highly diverse viral strain, subtype or group of an
CC endogenous retrovirus. Also included are an isolated ancestor protein (or
CC its fragment) from an endogenous retrovirus, an isolated expression
CC construct (comprising the following operably linked elements:- a
CC transcriptional promoter; a nucleic acid encoding an endogenous
CC retrovirus ancestor protein and a transcriptional terminator), a cultured
CC prokaryotic (or eukaryotic cell) transfected or transfected with the
CC expression construct, an isolated host cell comprising the expression
CC construct, a composition for inducing an immune response in a recipient
CC mammal (comprising a viral ancestor protein or its antigenic fragment,
CC where the protein is from a virus of a donor species), an isolated
CC antibody that binds specifically to an endogenous retrovirus ancestor
CC protein (and that binds specifically to a plurality of circulating
CC descendant endogenous retrovirus ancestor proteins), a method of
CC preparing an ancestral endogenous retroviral amino acid sequence, a
CC method for inducing an immune response to a donor virus in a transplant
CC recipient or a potential transplant recipient, a method of making a
CC vaccine, a method for detecting infection with an endogenous retrovirus
CC and a method for performing xenotransplantation in a subject. The
CC ancestral viral nucleic acid sequence is of Porcine Endogenous Retrovirus
CC (PERV) subtype A, B or C and has at least 70% identity with sequence of
CC any of the 18 nucleotide sequences fully defined in the specification.
CC The sequence may also be optimized for expression in a human host. The
CC nucleic acid sequence or its fragment is useful in the prophylaxis of
CC viral infection in transplantation that is heightened by the presence of
CC factors commonly associated with viral activation, e.g. immune
CC suppression, graft versus host disease, graft rejection, viral co-
CC infection, and cytotoxic therapies. The present sequence is an ancestral
CC PERV env protein.
XX
SQ Sequence 638 AA;
XX
Alignment Scores:
Pred. No.: 1,08e-289 Length: 638
Score: 3407.00 Matches: 635
Percent Similarity: 99.7% Conservative: 1
Best Local Similarity: 99.5% Mismatches: 2
Query Match: 98.2% Indels: 0
DB: Gaps: 0
US-10-723-552-3_COPY_5620_7533 (1-1914) x ADY28028 (1-638)
QY 1 ATGCATCCCGTAAACCCGCGCCACCTCCCGATTCCGGGTGGAAGCCGAAAGACTG 60
Db 1 MethisProThrLeuSerArgArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20
QY 61 AAAATCCCTTAAAGCTTCGCTCCATCGCGTGGTTCCTACTCTGTCAATACCTCTCAG 120
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40
QY 121 ACTAATGCTATGCCATAGGAGACAGCTGAACCTCCATAAACCTTATCTCTCACCTGG 180
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTACTGCTCCGGCACAGGTATTAAATATCAACAACACTCAAGGGAGGCTCTTTTA 240
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnThrGlnGlyGluAlaProLeu 80
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCTCAGATCAGTTATCTAGTCTGACCTCA 300
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100
QY 301 CCCCAGATATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACCAACAAATAATCGA 360
Db 101 ProProaspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120
QY 361 AAACATTCGCGAAATCCAGAGATTCTTTTGTAAACAATGGAACGTGTGAACCTTAAT 420
Db |||||||
```

Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140  
Qy 421 GATGATATTGGAATGCCCAACCTCTCAGCAGATAGGTAAGTTTTCTTATGTCAAC 480  
Db 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
Qy 481 ACCTATACAGCTCTGGACAATTTAATTACCTGACCTCGATTTAGAACTCGGAAGCCCAAG 540  
Db 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpLysArgThrGlySerProLys 180  
Qy 541 TGCTCTCTCAGACCTAGATTACCTAAATAAGTTTCACTGAGAAAGGAACACAGAA 600  
Db 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGln 200  
Qy 601 AATATCCATAAATGGTAAATGGTATGCTTTGGGAATGGTATATATGAGAGGCTCGGGT 660  
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGly 220  
Qy 661 AAACACACAGGCTCATTCTAATCTATTCGCTCAAAATAAACACAGCTGGAGCCCTCAATG 720  
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
Qy 721 GCTATAGGACCAATACCGTCTTCACGGGTCAAGAGCCCAAGACCCCAAGCAGGACCA 780  
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
Qy 781 TCCTCTAACATAACTCTCGATCAGACCCCACTGAGTCTAAGCAGCAGCACTAAATGGG 840  
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
Qy 841 GCAAACTTTTACGCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCACGACTCCAGAG 900  
Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
Qy 901 GCTACCTCTTCTTGGCTATGCTAGCTTCGGGCCCACTTACTATGAGGAATGGCT 960  
Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGlyGlyMetAla 320  
Qy 961 AGAAGAGGGAATTCATGTGACAAAGAACATAGAGACCAATGCATCGGGATCCCAA 1020  
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
Qy 1021 AATAAGCTTACCTTACGAGTTTCTGAAAGGACCTGCATAGAAAGGTTTCCCCCA 1080  
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
Qy 1081 TCCACCAACACCTTTTACCACTGACCTTAACTCAAGCTTAACTCAAGCTTCAAGCTCAAT 1140  
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
Qy 1141 CTGGTACCTGGTTATGACAGGTGGGCGCATGTAACTAGATTAAACCCCTTGTGTTC 1200  
Db 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
Qy 1201 ACCTGGTTTAAACCAAACTAAAGATTTTGCATTATGTTCCAAATGTTTCCCGAGTG 1260  
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
Qy 1261 TATTACTATCCGAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGCAAAAG 1320  
Db 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
Qy 1321 AGAAGACCCATATCTGACACTTGTGTGTATGCTCGGACTTGGAGTGGCAGCAGGTGA 1380  
Db 441 ArgGluProLysLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIleGlyVal 460  
Qy 1381 GGAACAGGAACAGCTGCCCTGGTCACGGGACCAAGCAGCAGCTAGAAAACAGGACTTAGTAAC 1440  
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
Qy 1441 CTACATCGAATTTGTAACAGAGATCTCCAGCCCTAGAAAATCTGTCAATACCTGGAG 1500  
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500

Qy 1501 GAAATCCCTAACCTCTTATCTGAAGTAGTCTCAGCAATAGAGGGTTAGATTATTA 1560  
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
Qy 1561 TTTCTTAAAGAGAGGAGATTATGTAGCTTGTAGCCCTTGAAGGAGGAATGCTGTTTTATGTGGAT 1620  
Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluCysCysPheTyrValAsp 540  
Qy 1621 CATTACGGGGCCATCAGAGACTCCATCAAGCAAGCTTAGAGAAAGGTTGGAGAGCGTCGA 1680  
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluLeuLysArgArg 560  
Qy 1681 AGGAAAGGAAACTACTCAAGGCTGTTGAGGAGTGGTTCAACAGGCTCTCTTTGGTTG 1740  
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580  
Qy 1741 GCTACCTACTTCTTCTTAAACAGGACCTTAATAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1800  
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrVal 600  
Qy 1801 GGGCCATGTATTATTAACAAGTTAATTCCTTCAATTAGAGAACGAATAGTCAGTCCAG 1860  
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
Qy 1861 ATCATGTACTTAGACAACTAGTACCAAGCCCGTCTAGCAGGGAAGCTGGCCGC 1914  
Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 638

## RESULT 7

ADY28030

ID ADY28030 standard; protein; 638 AA.

AC ADY28030;

XX 19-MAY-2005 (first entry)

XX Porcine endogenous retrovirus subtype C ancestral env protein, SEQ ID 34.  
XX Xenotransplantation; vaccine; acquired immune deficiency syndrome;  
XX anti-hiv; immune disorder; HIV-1 infection; infection; PERV infection;  
XX diagnosis; graft versus host disease; immunosuppressive;  
XX envelope protein.

OS Porcine endogenous retrovirus; subtype C.  
OS Synthetic.

XX WO2005019411-A2.

XX 03-MAR-2005.

XX 19-MAY-2004; 2004WO-US015709.

XX 19-MAY-2003; 2003US-00441949.

XX (AUCC-) AUCCLAND UNISERVICES LTD.  
XX (UNIW ) UNIV WASHINGTON.

XX Mullins JI, Rodrigo A, Ross HA;

XX WPI; 2005-202625/21.

XX N-PSDB; ADY28029.

XX New isolated ancestral viral nucleic acid sequence that is a determined  
XX founder sequence of a highly diverse viral strain, subtype or group of an  
XX endogenous retrovirus, useful in the prophylaxis of viral infection.

XX Claim 10; SEQ ID NO 34; 278pp; English.

XX The invention relates to an isolated ancestral viral nucleic acid  
XX sequence and its fragment, where the sequence is a determined founder  
XX sequence of a highly diverse viral strain, subtype or group of an  
XX endogenous retrovirus. Also included are an isolated ancestor protein (or

CC its fragment) from an endogenous retrovirus, an isolated expression  
CC construct (comprising the following operably linked elements:- a  
CC transcripional promoter; a nucleic acid encoding an endogenous  
CC retrovirus ancestor protein and a transcriptional terminator), a cultured  
CC prokaryotic (or eukaryotic cell) transformed or transfectd with the  
CC expression construct, an isolated host cell comprising the expression  
CC construct, a composition for inducing an immune response in a recipient  
CC mammal (comprising a viral ancestor protein or its antigenic fragment,  
CC where the protein is from a virus of a donor species), an isolated  
CC antibody that binds specifically to an endogenous retrovirus ancestor  
CC protein (and that binds specifically to a plurality of circulating  
CC descendant endogenous retrovirus ancestor proteins), a method of  
CC preparing an ancestral endogenous retroviral amino acid sequence, a  
CC method for inducing an immune response to a donor virus in a transplant  
CC recipient or a potential transplant recipient, a method of making a  
CC vaccine, a method for detecting infection with an endogenous retrovirus  
CC and a method for performing xenotransplantation in a subject. The  
CC ancestral viral nucleic acid sequence is of Porcine Endogenous Retrovirus  
CC (PERV) subtype A, B or C and has at least 70% identity with sequence of  
CC any of the 18 nucleotide sequences fully defined in the specification.  
CC The sequence may also be optimized for expression in a human host. The  
CC nucleic acid sequence or its fragment is useful in the prophylaxis of  
CC viral infection in transplantation that is heightened by the presence of  
CC factors commonly associated with viral activation, e.g. immune  
CC suppression, graft versus host disease, graft rejection, viral co-  
CC infection, and cytotoxic therapies. The present sequence is an ancestral  
CC PERV env protein.  
XX  
SQ Sequence 638 AA;

Alignment Scores:  
Pred. No.: Length: 638  
Score: 3407.00 Matches: 635  
Percent Similarity: 99.7% Conservative: 1  
Best Local Similarity: 99.5% Mismatches: 2  
Query Match: 98.2% Indels: 0  
DB: 9 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x AD28030 (1-638)

QY 1 ATGCATCCAGCTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAAGCCGAAAGACTG 60  
DB 1 MethHisProThrLeuSerArgHisLeuProIleArgGlyGlyLeuProLysArgLeu 20  
QY 61 AAAATCCCTTAAAGCTTCCTCCATCGCGGTGTTCTTACTCTGTCTCAATTAACCTCTCAG 120  
DB 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
QY 121 ACTAATGCTATCGCATAGGACAGCCTGAACTCCCATTAACCTTATCTCTCACCTGG 180  
DB 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 181 TTAATTACTGACTCCGGCACAGGTATTAAATATCAACACACTCAAGGGGAGCTCTTTA 240  
DB 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80  
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTTCAGATCAGTATTTCCTAGTCTGACCTCA 300  
DB 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
QY 301 CCCCAGATATCTCCAGTCTCACGGATTTATGTTTCCCGAGGACCCACCAATATGGA 360  
DB 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120  
QY 361 AAACATTCCGGAATCCACAGATTTCTTTTGTAAACAATGGAACCTGTAACTCTTAAT 420  
DB 121 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 140  
QY 421 GATGATATTTGAAATGGCCAACTCTCTCAGCAGGATAGGGTAAGTTTCTTATGTCAC 480  
DB 141 AspGlyTyrTrpLysTrpProThrSerGlnAspArgValSerPheSerTyrValAsn 160  
QY 481 ACCTATACGAGCTCTGGACAAATTTAATTACCTGACCTGGATTAGAACTGGAAGCCCAAG 540

DB 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180  
QY 541 TGCTCTCTCTCAGACCTAGATTACCTAAAAATAAGTTTCTACTGAGAAAGGAAACAAGAA 600  
DB 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
QY 601 AATATCTTAAATGGGTAAATGGTATGCTCTTGGGGAATGGTATATTATGGAGGCTCGGGT 660  
DB 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGly 220  
QY 661 AAACAACAGAGCTCCATTCTAACTATTTCGCTCAAAATAAACACGACTGGAGCTCCCAATG 720  
DB 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
QY 721 GCTATAGACCAATATACGGTCTTGACGGGTCAAAAGACCCCAACCAAGGACGAGGACCA 780  
DB 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
QY 781 TCCTCTAACATACTTCTGGATCAGACCCCACTGAGTCTAACACGACGACTAAATATGGG 840  
DB 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
QY 841 GCAAAACTTTTACCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAG 900  
DB 281 AlalysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 901 GCTACCTCTTCTTGGCTATGCTTAGCTTCGGGCCACCTTACTATGAAGGAATGGCT 960  
DB 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAla 320  
QY 961 AGAAGAGGGAATTCATGTGACAAAGAACATAGACCAATGACATGCGGATGCCAA 1020  
DB 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 1021 AATAAGCTTACCTTACTGAGGTTTCTGGAAAGGACCTGTCATAGGAAAGGTTTCCCCCA 1080  
DB 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 1081 TCCCAACAACACCTTTGTAAACACACTGAAGCCTTTTAACTCAACCTCTGAGAGTCAATAT 1140  
DB 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
QY 1141 CTGCTACTCTGGTATGACAGGTGGTGGCATGTAATACTGGATTAAACCCCTTGTTGTTTCC 1200  
DB 381 LeuValProGlyTyrAspArgTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 1201 ACCTTGGTTTTTAAACCAAACTTAAAGATTTTTCATTTATGGTCCAAATTTGTTCCCGAGTG 1260  
DB 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 1261 TATTACTATCCGAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAG 1320  
DB 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
QY 1321 AGACAACCATATCTCTGACACTTCTGTGATGCTCGGACTTGGGAGTGGCAGCAGGTGTA 1380  
DB 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
QY 1381 GGAACAGAACAGCTGCGCTCGTCACGGGACCAACAGACGCTAGAAAACAGGACTTAGTAAC 1440  
DB 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 1441 CTACATCGAATGTGACAGAGATCTCCAAGCCCTAGAAAAATCTGCAGTAACCTGGAG 1500  
DB 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 1501 GAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGATAGAGAGGTTAGATTATTA 1560  
DB 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeu 520  
QY 1561 TTTCTAAAGAGGAGGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTGGAT 1620







QY 601 AATATCCATAAATGGTAAATGGTATGCTTGGGAAATGGTATATTATGGAGGCTCGGCT 660  
Db |||||  
201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyTrGlyGlySerGly 220  
QY 661 AAACAACAGGCTCCATTCTAACTATTCGCTCAAAATAAAACCCAGCTGGAGCCCTCCCAATG 720  
Db |||||  
221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
QY 721 GCTATAGACCAATAATACGGCTTGACGGGTCAAGACCCCAACCCCAAGGACCCAGGACCA 780  
Db |||||  
241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
QY 781 TCCTCTACATAAATCTTGGAGTCAGACCCCACTAGCTTAACAGACAGCTAAATAATGGG 840  
Db |||||  
261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
QY 841 GCATAACCTTTTAGCCTCATCAGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAG 900  
Db |||||  
281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 901 GCTACCTCTCTTGTGGCTATGCTTACGCTTCGGGCCCACTTACTACTAAGAGGAATGGCT 960  
Db |||||  
301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyTrGlyGlyMetAla 320  
QY 961 AGAAGAGGGAATTCATGTGCAAAAGAACATAGAGACCAATGCACATGGGGATCCCAA 1020  
Db |||||  
321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 1021 AATAAGCTTACCTTACTGAGGTTTCTCGAAAAGGCACCTGCATAGGAAAGGTTTCCCCCA 1080  
Db |||||  
341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 1081 TCCCAACCAACCTTTGTAAACACACTGAAGCTTTAATCAAACTCTGAGAGTCAATAT 1140  
Db |||||  
361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTy 380  
QY 1141 CTGTGTACCTGTTATGACAGTGGTGGCATGTATACTGATTAACCCCTTGTGTTTCC 1200  
Db |||||  
381 LeuValProGlyTyTrpAspArgTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 1201 ACCTTGGTGTTTTAAACCAACTAAAGATTTTGCATTATGGTCCCAATTTGTTCCCGAGTG 1260  
Db |||||  
401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 1261 TATTAATATCCGGAAGAAAGCAATCTTGATGATATGACTACAGAAATCATCGACAAAG 1320  
Db |||||  
421 TyTrpTyTrpProGluLysAlaIleLeuAspGluTyTrpAspTyTrpArgAsnHisArgGlnLys 440  
QY 1321 AGAGAACCATATCTCTGACACTTCTGATGCTCGGACTTGGAGTGGCAGCAGGTGTA 1380  
Db |||||  
441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
QY 1381 GGAACAGGACAGCTCGCTGTCTACGGGACCAACAGCAGCTAGAACAGGACTTAGTAAC 1440  
Db |||||  
461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 1441 CTACATCAATGTAAACAGAGATCTCCAGCCCTAGAAAAATCTGTGAGTAACTGAG 1500  
Db |||||  
481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 1501 GAATCCCTTAACCTCTTATCTGAAGTGTCTTACAGATAGAGAGGGGTAGATTATTA 1560  
Db |||||  
501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
QY 1561 TTTCTAAAGAGGAGGATTTATGTAGCTTTGAAGGAGGAATGCTGTTTTATGTGGAT 1620  
Db |||||  
521 PheLeuLysGluGlyLeuCysValAlaLeuLysGluGluCysCysPheTyValAsp 540  
QY 1621 CATTCAGGGGCCATCAGAGACTCCATGAACAGCTTTAGAGAAAGTTGGAGAGGCTGCA 1680  
Db |||||  
541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluLeuGluLysArg 560  
QY 1681 AGGGAAGGAAACTACTCAAGGGTGGTTTGAGGGATGTTTCAACAGGCTCTCTTGGTTG 1740

Db |||||  
561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580  
QY 1741 GCTACCTACTTCTGCTTTAAACAGGACCCCTTAATAGTCTCTCTCTGTTACTCAGATT 1800  
Db |||||  
581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
QY 1801 GGCCCATGTATTATTAACAAAGTTAATTCCTCTTATTAGAGAACCAATAAGTGCAGTCCAG 1860  
Db |||||  
601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 1861 ATCATGTGTACTTACACACAGTACCAAGCCCGCTCTAGCAGGGAAGCTGGCCGC 1914  
Db |||||  
621 IleMetValLeuArgGlnGlnTyTrpGlnSerProSerSerArgGluAlaGlyArg 638  
RESULT 9  
ADY28041  
ID ADY28041 standard; protein; 638 AA.  
XX  
AC ADY28041;  
XX  
DT 19-MAY-2005 (first entry)  
XX  
DE Porcine endogenous retrovirus subtype C ancestral env protein, SEQ ID 45.  
XX  
KW Xenotransplantation; vaccine; acquired immune deficiency syndrome;  
KW anti-hiv; immune disorder; HIV-1 infection; infection; PERV infection;  
KW diagnosis; graft versus host disease; immunosuppressive;  
KW envelope protein.  
XX  
OS Porcine endogenous retrovirus; subtype C.  
OS Synthetic.  
XX  
FN WO2005019411-A2.  
XX  
PD 03-MAR-2005.  
XX  
PF 19-MAY-2004; 2004WO-US015709.  
XX  
PR 19-MAY-2003; 2003US-00441949.  
XX  
PA (AUCK-) AUCKLAND UNISERVICES LTD.  
XX (UNIW ) UNIV WASHINGTON.  
XX  
PI Mullins JJ, Rodrigo A, Ross HA;  
XX  
DR WPI; 2005-202625/21.  
XX  
PT New isolated ancestral viral nucleic acid sequence that is a determined  
PT founder sequence of a highly diverse viral strain, subtype or group of an  
PT endogenous retrovirus, useful in the prophylaxis of viral infection.  
XX  
FS Claim 10; SEQ ID NO 45; 278pp; English.  
XX  
CC The invention relates to an isolated ancestral viral nucleic acid  
CC sequence and its fragment, where the sequence is a determined founder  
CC sequence of a highly diverse viral strain, subtype or group of an  
CC endogenous retrovirus. Also included are an isolated ancestor protein (or  
CC its fragment) from an endogenous retrovirus, an isolated expression  
CC construct (comprising the following operably linked elements:- a  
CC transcriptional promoter; a nucleic acid encoding an endogenous  
CC retrovirus ancestor protein and a transcriptional terminator), a cultured  
CC prokaryotic (or eukaryotic cell) transformed or transfected with the  
CC expression construct, an isolated host cell comprising the expression  
CC construct, a composition for inducing an immune response in a recipient  
CC mammal (comprising a viral ancestor protein or its antigenic fragment,  
CC where the protein is from a virus of a donor species), an isolated  
CC antibody that binds specifically to an endogenous retrovirus ancestor  
CC protein (and that binds specifically to a plurality of circulating  
CC descendant endogenous retrovirus ancestor proteins), a method of  
CC preparing an ancestral endogenous retroviral amino acid sequence, a  
CC method for inducing an immune response to a donor virus in a transplant  
CC recipient or a potential transplant recipient, a method of making a

CC vaccine, a method for detecting infection with an endogenous retrovirus  
 CC and a method for performing xenotransplantation in a subject. The  
 CC ancestral viral nucleic acid sequence is of Porcine Endogenous Retrovirus  
 CC (PERV) subtype A, B or C and has at least 70% identity with sequence of  
 CC any of the 18 nucleotide sequences fully defined in the specification.  
 CC The sequence may also be optimized for expression in a human host. The  
 CC nucleic acid sequence or its fragment is useful in the prophylaxis of  
 CC viral infection in transplantation that is heightened by the presence of  
 CC factors commonly associated with viral activation, e.g. immune  
 CC suppression, graft versus host disease, graft rejection, viral co-  
 CC infection, and cytotoxic therapies. The present sequence is an ancestral  
 CC PERV env protein.

XX  
 SQ Sequence 638 AA;

## Alignment Scores:

Pred. No.: 1.06e-289 Length: 638  
 Score: 3407.00 Matches: 635  
 Percent Similarity: 99.7% Conservative: 1  
 Best Local Similarity: 99.5% Mismatches: 2  
 Query Match: 98.2% Indels: 0  
 DB: 9 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x ADY28041 (1-638)

QY 1 ATGCATCCACGTTAAACCGCGCCACCTCCCGATTCGGGTGGAAGCCGAAAGACTG 60  
 DB : : : : :  
 QY 1 MethisProThrLeuSerArgArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20  
 61 AAAATCCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCATTAACCTCTCAG 120  
 DB : : : : :  
 QY 21 LysileProLeuSerPheAlaSerIleAlaIlePheLeuThrLeuSerIleThrSerGln 40  
 121 ACTAATGGTATGCCATAGGAGACAGCTGAACCTCCCAATCAACCTTATCTCTCACCTGG 180  
 DB : : : : :  
 QY 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
 181 TTAATTACTGACTCCGCGACAGGTATTAAATATCAACCAACTCAAGGGGAGGCTCCTTTA 240  
 DB : : : : :  
 QY 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnLeuThrGlnGlyGluAlaProLeu 80  
 241 GGAACCTGGTGGCTGATCTATAGTTTGGCTCGATCAGTATTCTCTAGCTGACCTCA 300  
 DB : : : : :  
 QY 81 GlyThrTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
 301 CCCCAGATATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACCAACCAATAATGGA 360  
 DB : : : : :  
 QY 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120  
 361 AAACATTGCGGAATCCAGAGATTTCTTTGTAACAATGGAACCTGTGTAACCTCTAAT 420  
 DB : : : : :  
 QY 121 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 140  
 421 GATGGATATTGGAATGCCAACCTCTCAGCAGATAGGTAAGTTTCTTATGTCAAC 480  
 DB : : : : :  
 QY 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
 481 ACCTATACCACTCTGACACATTTAATTTACTGACCTGGATTAGAACCTGGAACCCCAAG 540  
 DB : : : : :  
 QY 161 ThrTyrThrSerSerGlyGlnPheAsnTrpLeuThrTrpIleArgThrGlySerProLys 180  
 541 TGCTCTCTCTCAGACCTAGATTACTAAAAATAGTTTCACTGAGAAGGAAACCAAGAA 600  
 DB : : : : :  
 QY 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
 601 AATATCTAAATGGGTAATGGTATGCTCTTGGGAATGGTATATTATGAGGCTCGGGT 660  
 DB : : : : :  
 QY 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGly 220  
 661 AAACACCAAGCTCCATTTAACTATTCGCCTCAAAATAACCAAGCTGGAGCTCCCAATG 720  
 DB : : : : :  
 QY 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240

QY 721 GCTATAGACCAAAATACGGTCTTGACGGGTCAAAAGACCCCAACCCCAAGAGCAGGACCA 780  
 DB : : : : :  
 QY 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
 781 TCCTCTAACATAACTCTCGATCAGCCCACTCAGTCTAACACACGACTAAATATGGG 840  
 DB : : : : :  
 QY 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
 841 GCAAAACCTTTTACCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCACGACTCCAGAG 900  
 DB : : : : :  
 QY 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
 901 GCTACCTCTCTTCTGCTATGCTTAGCTTCGGGCCCCACCTTACTATGAAGAAATGCT 960  
 DB : : : : :  
 QY 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAla 320  
 961 AGAAGAGGGAAATTCATGTGACAAAGACATAGACCAATCAACATGGGGATCCCAA 1020  
 DB : : : : :  
 QY 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
 1021 AATAAGCTTACCTTACTGAGGTTTCTGGAAAAGGCACCTGCATAGGAAAAGTTTCCCCCA 1080  
 DB : : : : :  
 QY 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
 1081 TCCCAACCAACCTTTGTAACCACTGAAGCCTTTAATCAAACTCTGAGAGTCAATAT 1140  
 DB : : : : :  
 QY 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
 1141 CTGGTACTCTGGTATGACAGGTGGTGGCATGTAATACTGGATTAAACCTTGTGTTC 1200  
 DB : : : : :  
 QY 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
 1201 ACCTTGGTGTAAACCAAACTTAAAGATTTTGGCATTTATGGTCCAAATTTGTTCCCGAGTG 1260  
 DB : : : : :  
 QY 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
 1261 TATTACTATCCGAAAAAGCAATCTTTGATGAATATCACTACAGAAATCATCGACAAAG 1320  
 DB : : : : :  
 QY 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
 1321 AGAAGACCATATCTCTGACACTTGTGTCATGCTCGGACTTGGAGTGGGACGAGGTGTA 1380  
 DB : : : : :  
 QY 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIleGlyVal 460  
 1381 GGAACAGGAACAGCTGCCCTGGTCACGGGACCAAGCAGCTAGAAAAACAGACTTAGTAAC 1440  
 DB : : : : :  
 QY 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
 1441 CTACATCGAATTTGTAACAGAAATCTCCAAGCCCTAGAAAAATCTGTCACTAACCTGGAG 1500  
 DB : : : : :  
 QY 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
 1501 GAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGATAGAGAGGTTAGATTATTATTA 1560  
 DB : : : : :  
 QY 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
 1561 TTTCTAAAACAGAGGATTTATGTAGCTCTGAAGGAGGAATGCTGCTTTTATGTGCGAT 1620  
 DB : : : : :  
 QY 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluCysPheTyrValAsp 540  
 1621 CATTACGGGGCCCATCAGAGACTCCATGAAACAACTTAGAGAAAGTTGGAGAGCGTCGA 1680  
 DB : : : : :  
 QY 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560  
 1681 AGGAAAAGAAACTACTCAAGGGTGGTTGAGGGATGGTTCAACAGGTCTCTTTGGTTG 1740  
 DB : : : : :  
 QY 561 ArgGluLysGluThrThrGlnGlyTyrPheGluGlyTrpPheAsnArgSerProTrpLeu 580  
 1741 GCTACCTCTCTCTTAAACAGGACCTTAATAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1800  
 DB : : : : :  
 QY 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrVal 600  
 1801 GGGCCATGTATTATTAAACAAAGTTAATTTGCTTTCATTAGAGAACGAATAAGTCAGTCCAG 1860

Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 1861 ATCATGGTACTAGACAACAGTACCAAGCCGCTAGCAGGAGAGCTGGCGC 1914  
Db 621 IleMetValLeuArgGlnGlnTrpGlnSerProSerArgGluAlaGlyArg 638

RESULT 10  
ADY28038  
ID ADY28038 standard; protein; 638 AA.

XX AC ADY28038;  
XX 19-MAY-2005 (first entry)  
XX

DE Porcine endogenous retrovirus subtype C ancestral env protein, SEQ ID 42.  
KW Xenotransplantation; vaccine; acquired immune deficiency syndrome;  
KW anti-hiv; immune disorder; HIV-1 infection; infection; PERV infection;  
KW diagnosis; graft versus host disease; immunosuppressive;  
KW envelope protein.

OS Porcine endogenous retrovirus; subtype C.  
OS Synthetic.  
XX WO2005019411-A2.

XX PD 03-MAR-2005.  
XX PF 19-MAY-2004; 2004WO-US015709.

XX PR 19-MAY-2003; 2003US-00441949.  
XX PA (AUCK-) AUCKLAND UNITSERVICES LTD.  
XX PA (UNITW) UNIV WASHINGTON.

XX PI Mullins JI, Rodrigo A, Ross HA;  
XX WPI; 2005-202625/21.  
DR N-PSDB; ADY28037.

XX New isolated ancestral viral nucleic acid sequence that is a determined  
PT founder sequence of a highly diverse viral strain, subtype or group of an  
PT endogenous retrovirus, useful in the prophylaxis of viral infection.

XX Claim 10; SEQ ID NO 42; 278pp; English.

XX The invention relates to an isolated ancestral viral nucleic acid  
CC sequence and its fragment, where the sequence is a determined founder  
CC sequence of a highly diverse viral strain, subtype or group of an  
CC endogenous retrovirus. Also included are an isolated ancestor protein (or  
CC its fragment) from an endogenous retrovirus, an isolated expression  
CC construct (comprising the following operably linked elements:- a  
CC transcriptional promoter; a nucleic acid encoding an endogenous  
CC retrovirus ancestor protein and a transcriptional terminator), a cultured  
CC prokaryotic (or eukaryotic cell) transformed or transfected with the  
CC expression construct, an isolated host cell comprising the expression  
CC construct, a composition for inducing an immune response in a recipient  
CC mammal (comprising a viral ancestor protein or its antigenic fragment,  
CC where the protein is from a virus of a donor species), an isolated  
CC antibody that binds specifically to an endogenous retrovirus ancestor  
CC protein (and that binds specifically to a plurality of circulating  
CC descendant endogenous retrovirus ancestor proteins), a method of  
CC preparing an ancestral endogenous retroviral amino acid sequence, a  
CC method for inducing an immune response to a donor virus in a transplant  
CC recipient or a potential transplant recipient, a method of making a  
CC vaccine, a method for detecting infection with an endogenous retrovirus  
CC and a method for performing xenotransplantation in a subject. The  
CC ancestral viral nucleic acid sequence is of Porcine Endogenous Retrovirus  
CC (PERV) subtype A, B or C and has at least 70% identity with sequence of  
CC any of the 18 nucleotide sequences fully defined in the specification.  
CC The sequence may also be optimized for expression in a human host. The  
CC nucleic acid sequence or its fragment is useful in the prophylaxis of

CC viral infection in transplantation that is heightened by the presence of  
CC factors commonly associated with viral activation, e.g. immune  
CC suppression, graft versus host disease, graft rejection, viral co-  
CC infection, and cytotoxic therapies. The present sequence is an ancestral  
XX PERV env protein.  
SQ Sequence 638 AA;

Alignment Scores:  
Pred. NO.: 1.06e-289 Length: 638  
Score: 3407.00 Matches: 635  
Percent Similarity: 99.7% Conservative: 1  
Best Local Similarity: 99.5% Mismatches: 2  
Query Match: 98.2% Indels: 0  
DB: 9 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x ADY28038 (1-638)

QY 1 ATGCATCCACGTTAAACCGCGGCACCTCCCGGATTCGGGTGGAAAGCCGAAAGACTG 60  
Db 1 MethisProThrLeuSerArgArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20  
QY 61 AAAATCCCTTAAAGCTTCGCCTCCATCGCGGTTCCTTACTCTGTCAATTAACCTCTCAG 120  
Db 21 LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
QY 121 ACTAATGGTATGCGCATAGGAGACAGCTGAACTCCCATAAACCTTATCTCACCTGG 180  
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 181 TTAATTACTGCTCCGCGCACAGAGTATTAAATCAACAACACTCAAGGCGGCTCTTTA 240  
Db 61 LeulleThrAspSerGlyThrGlyIleAsnIleAsnThrGlnGlyGluAlaProleu 80  
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTCAGATCAGTATTTCCTAGTCTGACCTCA 300  
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
QY 301 CCCCAGATATCTCCATGCTCAGGATTTTATGTTTCCCGGAGCCAGCCAAATAATGGA 360  
Db 101 ProProAspIleLeuHisAlaHisGlyPheTyValCysProGlyProProAsnAsnGly 120  
QY 361 AAACATTCGCGAAATCCAGAGATTCTTTTGTAAACAATCGAACTGTGTAACTCTTAAT 420  
Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140  
QY 421 GATCGATATTGGAATGCCAACCTCTCAGCAGATAGGTAAGTTTTTCTTATGTCTAAC 480  
Db 141 AspGlyTyTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyValAsn 160  
QY 481 ACCTATACCAGCTCTGGACATTTAATTACCTGACCTGGATTAGAACTGGAGCCCAAG 540  
Db 161 ThrTyThrSerSerGlyGlnPheAsnTyLeuThrTrpIleArgThrGlySerProLys 180  
QY 541 TGCTCTCTCCAGACCTAGATTACTTAAATAATAAGTTTCACTGAGAAAGGAAAAACAAGA 600  
Db 181 CysSerProSerAspLeuAspTyLeuLysIleSerPheThrGluLysGlyGlnGlu 200  
QY 601 AATATCTCTAAATGGGTAATGGTATGTTCTTGGGGAATGGTATATATATGAGGCTCGGGT 660  
Db 201 AsnileLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyTrGlyGlySerGly 220  
QY 661 AAACACACGCTCCATCTTAACCTATTTCGCTCAAAATAAACCCAGCTGGAGCCTCCAATG 720  
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
QY 721 GCTATAGACCAATAATACGGTCTTCAGCGGTCAAAGACCCCAACCCAGGACCAGGACCA 780  
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
QY 781 TCCTCTAACATACTTCTGGATCAGACCCCACTGAGTCTAACAGCACGACTAAATATGGG 840  
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280

QY 841 GCAAACTTTTGTAGCTCATCCAGGAGCTTTTCAAGCTTTAACTCCAGACTCCAGAG 900  
Db 281 AlalysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 901 GCTACCTCTTCTTGTGGCTATGCTAGCTTCGGGCCACCTTACTATGAAGGAATGGCT 960  
Db 301 AlalysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 320  
QY 961 AGAAGAGGAAATCAATGTGACAAAGAACATAGAGACCAATGCACATGGGGATCCCAA 1020  
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 1021 AATAAGCTTACCTTGTAGGTTTCTGGAAGGACCTGCATAGAGAAAGTTTCCCA 1080  
Db 341 AsnLysLeuThrLeuGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 1081 TCCACCAACACCTTGTAAACACACCTGAAGCCTTAACTCAACCTCTGAGAGTCAATAT 1140  
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
QY 1141 CTGTTACCTGTTATGACAGCTGGTGGCATGTAATCTGATTAAACCCCTTGTGTTCC 1200  
Db 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 1201 ACCTTGTTTAAACCAAACTAAAGATTTTTCGATTATGTTCCAAATGTTTCCCGAGTG 1260  
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 1261 TATTACTATCCGAAAGCAATCTTGATCAATATGACTACAGAAATCATGCAAAAG 1320  
Db 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
QY 1321 AGAAGACCATATCTGACACTTGTCTGTATGCTCGGACTTGGAGTGGCAGCAGGTGA 1380  
Db 441 ArgGluProLysLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIleValVal 460  
QY 1381 GGAAACAGGACAGCTGCCCTGGTCAACGGGACACAGCAGCTAGAAAACAGGACTTAGTAAC 1440  
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 1441 CTACATCGAATTGTAACAGAGATCTCCAGCCCTAGAAAATCTGCAGTAACCTGGAG 1500  
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 1501 GAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGAAATCAAGAGGTTAGATTATTA 1560  
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
QY 1561 TTTCTAAAAGAGGAGGATTATGTGTAGCCTTGAAGGAGGAATCTGTTTTTATGTGGAT 1620  
Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysPheTyrValAsp 540  
QY 1621 CATTCAGGGCCATCAGAGACTCATGAAACAGCTTAGAAAAGGTTGGAGAGCTGCA 1680  
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluLeuLysArgArg 560  
QY 1681 AGGAAAAGGAACTACTCAAGGGTGGTTTGAGGGATGGTTCAACAGCTCTCTTTGGTGG 1740  
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580  
QY 1741 GCTACCTTACTTCTGCTTTAAACAGGACCTTAAATAGTCTCTCTCTCTTACTCACAGTT 1800  
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
QY 1801 GGGCCATGTTATTAAAGATTAAATGCTTCAATAGAGAACGAATAAGTGCAGTCCAG 1860  
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 1861 ATCATGTACTTAGACACAGTACCAAGCCGCTAGCAGGAGCTGGCCGC 1914  
Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg 638

## RESULT 11

AAB70634

ID AAB70634 standard; protein; 638 AA.

XX AAB70634;

AC AAB70634;

DT 15-MAY-2001 (first entry)

DE Porcine endogenous retrovirus envelope protein TSUKUBA.

KW Porcine endogenous retrovirus; PoERV; gag; env; detection; antigenic; immunogenic; antiviral; vaccine; antiserum; viral infection.

OS Porcine endogenous retrovirus.

PN WO200112816-A2.

PD 22-FEB-2001.

PF 16-AUG-2000; 2000WO-GB003159.

PR 18-AUG-1999; 99GB-00019604.

XX (QUIP-) QUIP TECHNOLOGY LTD.

XX Galbraith D, Kelly H, Smith K;

XX WPI; 2001-211223/21.

XX New porcine endogenous retrovirus polypeptide fragment with retroviral specific antigenic or immunogenic activity, for detection of retroviral antibodies in a sample and in therapy or diagnosis.

PS Claim 7; Fig 2; 43pp; English.

XX The present invention describes a porcine endogenous retrovirus (PoERV) polypeptide fragment (I) which has PoERV specific antigenic or immunogenic activity. Also described are: (1) an antiserum (II) specific to (I); (2) a PoERV specific antibody (III) or its fragment raised against (I); (3) use of (I) in detecting (III) in a sample; (4) use of (III) in detecting PoERV in a sample; (5) use of (I) or (III) in therapy or diagnosis; and (6) an assay kit (IV) comprising (I) or (III) for use in detection of PoERV in a sample. (I) has antiviral activity, and can be used in vaccine production. (I) is useful in the detection of PoERV antibodies (III) in a sample and (III) is useful in the detection of PoERV in a sample. (I) and (III) are useful in therapy or diagnosis. Antiserum (II) to (I) and kits (IV) comprising (I) or (III) are useful in detection of specific PoERV types. Viral infection can be monitored and by using (I) or (III). The present sequence represents a specifically claimed PoERV envelope (env) protein sequence from the present invention

XX Sequence 638 AA;

## Alignment Scores:

Pred. No.:	2,92e-289	Length:	638
Score:	3402.00	Matches:	634
Percent Similarity:	99.7%	Conservative:	2
Best Local Similarity:	99.4%	Mismatches:	2
Query Match:	98.1%	Indels:	0
DB:	4	Gaps:	0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x AAB70634 (1-638)

QY	1	ATCATCCCAAGCTTAAACGGCGCCACCTCCCGATTCGGGTGGAAGCCGAAAGACTG	60
Db	1	MetHisProThrLeuSerArgHisLeuProIleArgGlyLysProLysArgLeu	20
QY	61	AAAATCCCTTAAGCTTCGCTCATCGCTGGTTCCTTACTCTGTCAATAACCTCTCAG	120
Db	21	LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln	40
QY	121	ACTAATGGTATGCGCATAGGAGACGCTGAACCTCCCAATAACCTTATCTCTCACCTGG	180

```
Db      41  ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp  60
Qy      181  TTAATTACTGACTCGGCACAGGTATTAAATATCAACACTCAAGGGAGGCTCCITTA  240
Db      61  LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyIleAlaProLeu  80
Qy     241  GGAACCTGGTGGCTGTACTTACCTTTGCTTCAGATCAGTATTATTCCTAGTCTCACCTCA  300
Db      81  GlyThrTrpTrpProAspLeuThrValCysLeuArgSerValIleProSerLeuThrSer  100
Qy     301  CCCCCAGATATCCTCCATCTCACGGATTTTATGTTTCCCGAGGACCAACCAATAATGGA  360
Db     101  ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly  120
Qy     361  AAACATTCGGGAATCCAGAGATTCTTTTGTGTAACAATGGAACGTGTGAACCTCTAAT  420
Db     121  LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn  140
Qy     421  GATGGATATTGGAATGGCCAACTCTCAGCAGGATAGGGTAAGTTTCTTATGTCTCAAC  480
Db     141  AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn  160
Qy     481  ACCTATACCAGCTCTGGCAATTTAAATTACCTGACCTGGATTAGAACTGGAGCCCAAG  540
Db     161  ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys  180
Qy     541  TGCTCTCTCAGACCTAGATTACCTAAATAAGTTTCTACTGTGAGAAAGGAAACAAGAA  600
Db     181  CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu  200
Qy     601  AATATCTCTAAATGGTAATGGTATGCTTGGGAATGGTATATTATGGAGGCTCGGGT  660
Db     201  AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyLysSerGly  220
Qy     661  AAACAACCAAGGCTCCATCTCACTATTCGCTCAAAATAAACCAAGCTGGAGCTCCCAATG  720
Db     221  LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProMet  240
Qy     721  GCTATAGACCAATACGGCTTGTAGCGGTGTAAGACCCCAACCCCAAGGACCGAGCA  780
Db     241  AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro  260
Qy     781  TCCTCTAACAATACTTCCTGGATCAGACCCCACTAGTCTTAACAGCACCACTAAATAATGGG  840
Db     261  SerSerAsnIleThrSerGlySerAspProThrGluSerSerSerThrThrLysMetGly  280
Qy     841  GCAAAACCTTTTACGCTCATCCAGGGAGCTTTTCAAGCTCTTAACCTCCACCACTCCAGAG  900
Db     281  AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu  300
Qy     901  GCTACTCTTCTTGTGGTATGCTTAGCTTCGGGCCCACTTACTACTATGAAGGAATGGCT  960
Db     301  AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAla  320
Qy     961  AGACAGGGAATTCAAATGTGAAAAGAACATAGACCAATGCATGCGATGGGATCCCAA  1020
Db     321  ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln  340
Qy    1021  AATAAGCTTACCTTACTGAGGTTTCTGGAAGGCACCTGCATAGGAAAGGTTCCCCCA  1080
Db     341  AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro  360
Qy    1081  TCCCAACCAACACTTTGTAAACCACTGAAGCCTTTAATCAAACTCTGAGAGTCAATAT  1140
Db     361  SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr  380
Qy    1141  CTGGTACTCGGTATGACAGGTGGGGCATCTAATCTTGGATTAAACCCCTTGTCTTCC  1200
Db     381  LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer  400
Qy    1201  ACCTTGGTTTTTAAACCAACTAAAGATTTTTCATTATGGTCCAAATGTTTCCCGAGTG  1260
Db     401  ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal  420
```

```
Qy     1261  TATTACTATCCGAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGCAAAAAG  1320
Db     421  TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys  440
Qy    1321  AGAGAACCCATATCTCTGACACTTGTCTGTGATCTCGGACTTGGAGTGGCAGCAGGTGA  1380
Db     441  ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal  460
Qy    1381  GGAACAGAAACAGCTGCCCTGGTCCAGGGACCAACAGACAGCTAGAAAACAGGACTTAGTAA  1440
Db     461  GlyThrGlyThrAlaAlaLeuValThrGlyProGlnLeuGluThrGlyLeuSerAsn  480
Qy    1441  CTACATCGAATTGTAAACAGAAAGATCTCCAAGCCCTAGAAAAAATCTGTCAGTAACCTGGAG  1500
Db     481  LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu  500
Qy    1501  GAAATCCCTAACTCCTTATCTGAAGTAGTCTCAGCAATAGAAAGGGTTAGATTTATTA  1560
Db     501  GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu  520
Qy    1561  TTTCTAAAGAGAGAGATTATGTAGCTTGTAGCCTTGAAGGAGGAATGCTGTTTTTATGTGGAT  1620
Db     521  PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAsp  540
Qy    1621  CATTTCAGGGCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGCTTGGAGAAGCGTCGA  1680
Db     541  HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg  560
Qy    1681  AGGAAAAAGGAAACTACTCAAGGGTGGTTTGGAGGATGGTTTCAACAGGTCTCTTTGGTTG  1740
Db     561  ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu  580
Qy    1741  GCTACCTTACTTTCTGCTTTAAACAGGACCCCTTAATAGTCTCTCTCTCTGTTACTCACAGTT  1800
Db     581  AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal  600
Qy    1801  GGGCCATCTATTATTAAACAGTTAATTCCTTATTAGAGAACGAATAAGTGCAGTCCAG  1860
Db     601  GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln  620
Qy    1861  ATCATGTGTACTTAGACAACAGTACCAAGCCCGTCTAGCAGGGAAGCTGGCCGC  1914
Db     621  IleMetValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg  638
```

## RESULT 12

ADS73449

ID ADS73449 standard; protein; 666 AA.

XX AC ADS73449;

XX DT 16-DEC-2004 (first entry)

XX DE Swine retroviral (Teskuba-1) env protein.

XX KW Swine retroviral protein; immunosuppressive; gene therapy; env protein.

XX OS Porcine endogenous retrovirus.

XX FH Key Location/Qualifiers

XX FT Misc-difference 13 /note= "Encoded by TGA"

XX FT Misc-difference 19 /note= "Encoded by TGA"

XX PN US2004185435-A1.

XX PD 23-SEP-2004.

XX PF 26-NOV-2003; 2003US-00723552.

XX PR 14-DEC-1995; 95US-00572645.

XX PR 13-DEC-1996; 96US-00766528.

PR 14-SEP-2000; 2000US-00661858.  
XX (GEHO ) GEN HOSPITAL CORP.  
XX Fishman JA;  
XX WPI; 2004-689179/67.  
XX N-PSDB; ADS73367.  
XX  
PT New porcine retroviral polypeptide encoded by a nucleic acid, useful in  
PT evaluating an immunosuppressive treatment for the ability to activate a  
PT retrovirus, such as an endogenous porcine retrovirus.  
XX  
XX Claim 8; Fig 1; 83pp; English.  
XX  
XX The present invention relates to the swine retroviral polypeptides and  
XX their encoding nucleic acids. The methods and compositions of the present  
XX invention are useful for screening a cell or tissue, e.g. a heart, lung,  
XX liver, bone marrow, kidney, brain cells, neural tissue, pancreas and  
XX intestinal tissue xenograft, for the presence or expression of a swine or  
XX miniature swine retrovirus or retroviral sequence. The invention is also  
XX useful in evaluating an immunosuppressive treatment for the ability to  
XX activate a retrovirus such as an endogenous porcine retrovirus. The  
XX invention is also useful in gene therapy. The present sequence is the  
XX swine retroviral (Tsukuba-1) env protein.  
SQ Sequence 666 AA;  
  
Alignment Scores:  
Pred. No.: 2,97e-289 Length: 666  
Score: 3402.00 Matches: 634  
Percent Similarity: 99.7% Conservative: 2  
Best Local Similarity: 99.4% Mismatches: 2  
Query Match: 98.1% Indels: 0  
DB: 8 Gaps: 0  
  
US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x ADS73449 (1-666)  
  
QY 1 ATGCATCCACGTTAAACCGCGCCACCTCCCGATTCGGGTGAAAGCGAAAGACTG 60  
DB 29 MetHisProThrLeuSerArgHisLeuProIleArgGlyLysProLysArgLeu 48  
  
QY 61 AAAATCCCTTACGCTCGCTCATCGGTGGTTCCTTACTCTGTCAATAACCTCTCAG 120  
DB 49 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 68  
  
QY 121 ACTAATCGTATGCGCATAGAGACAGCTGAACCTCCCATAAACCTTATCTCAGCTGG 180  
DB 69 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 88  
  
QY 181 TTAATTACTGACTCCGCGACAGGTATTAAATATCAACAACTCAAGGGGAGGCTCCTTTA 240  
DB 89 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyAlaProLeu 108  
  
QY 241 GGAACCTGGTGGCTGATCTATAGTTTGCTCAGATCAGTTATTCCTAGTCTGACCTCA 300  
DB 109 GlyThrTrpTrpProAspLeuTrpValCysLeuArgSerValIleProSerLeuThrSer 128  
  
QY 301 CCCCCAGATATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACCAACAATAATGGA 360  
DB 129 ProProAspIleLeuHisAlaHisGlyPheTrpValCysProGlyProProAsnAsnGly 148  
  
QY 361 AAACATTGCGGAATCCAGAGATTTCTTTTGTAAACAATGGAACCTGTAACTCTTAAT 420  
DB 149 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 168  
  
QY 421 GATGGATATTGGAAATGCCAACCTCTCAGCAGATAGGGTAAGTTTCTTATGTCAAC 480  
DB 169 AspGlyTrpTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTrpValAsn 188  
  
QY 481 ACCTATACGAGCTCTGGACAAATTAATTACTGACCTGGATTAGAACTGGAAGCCCCAAG 540  
DB 189 ThrTrpThrSerSerGlyGlnPheAsnTrpLeuThrTrpIleArgThrGlySerProLys 208

QY 541 TGCTCTCTCTCAGACCTAGATTACTAAAAATAAGTTTCACTGAGAAGAGAAAAACAAGAA 600  
DB 209 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 228  
  
QY 601 AATATCCATAAATGGGTAATGGTATGTCTTTGGGAATGGTATATTATGGAGGCTCGGGT 660  
DB 229 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTrpGlyLysGly 248  
  
QY 661 AAACAACAGGCTCATTTCTTAACATTTCGCTCAAAATAAACACAGCTGGAGCCCTCCATG 720  
DB 249 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 268  
  
QY 721 GCTATAGGACCAATACGCTCTTCAGCGGTCAAAAGACCCCCCAACCCCAAGGACGAGCA 780  
DB 269 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 288  
  
QY 781 TCCTCTTAACATACTCTCGATCAGACCCCACTGAGTCTTAACACAGCAGACTAAAATGGG 840  
DB 289 SerSerAsnIleThrSerGlySerAspProThrGluSerSerSerThrThrLysMetGly 308  
  
QY 841 GCAAAACCTTTTACGCTCATCCAGGGAGCTTTCAAGCTCTTAACCTCCAGCAGCTCCAGAG 900  
DB 309 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 328  
  
QY 901 GCTACCTCTTCTTGTGGCTATGCTTAGCTTCGGGCCACCTTACTACTATGAAGGAATGGCT 960  
DB 329 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTrpGlyGlyMetAla 348  
  
QY 961 AGAAGAGGAAATTCATGTGACAAAAGAACATAGAGACCAATGCACATGGGGATCCCAA 1020  
DB 349 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 368  
  
QY 1021 AATAAGCTTACCTTACTGAGGTTCTGGAAGGACCTGCATAGAAAGGTTCCCCCA 1080  
DB 369 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 388  
  
QY 1081 TCCCAACCAACACCTTTGTAAACCACTGAAGCCCTTTAATCAAACTCTGAGAGTCAATAT 1140  
DB 389 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 408  
  
QY 1141 CTGGTACCTGGTTATGACAGGTGGTGGCATGTAATACTGGATTAAACCTTGTGTTC 1200  
DB 409 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 428  
  
QY 1201 ACCTTGGTTTTTAAACAAACTTAAAGATTTTTCATTTATGTCCTCCAAATGTTCCCGAGTG 1260  
DB 429 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 448  
  
QY 1261 TATTACTATCCCGAAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAG 1320  
DB 449 TyrTrpTrpProGlyLysAlaIleLeuAspGluTyrAspTrpArgAsnHisArgGlnLys 468  
  
QY 1321 AGAAGAACCCATATCTCTGACACTTGTGTGATGCTCGACTTCGGAGTGGAGGAGTGTA 1380  
DB 469 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 488  
  
QY 1381 GGAACAGGAACAGCTCCCTGGTCACGGGACCAAGCAGCAGCTAGAAACAGACACTTAGTAAC 1440  
DB 489 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 508  
  
QY 1441 CTACATCGAATTCTTAACAGAAAGTCTCAAGCCCTAGAAAAATCTGTCACTAACCTCGAG 1500  
DB 509 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGlyLysSerValSerAsnLeuGlu 528  
  
QY 1501 GAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGAATAAGAGAGGGTTAGATTATTA 1560  
DB 529 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 548  
  
QY 1561 TTCTTAAAGAGGAGGATTTATGTGTAGCTTGAAGGAGGAATGCTGTTTTTATGTGGAT 1620  
DB 549 PheLeuLysGlyGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAsp 568









CC where the protein is from a virus of a donor species), an isolated  
 CC antibody that binds specifically to an endogenous retrovirus ancestor  
 CC protein (and that binds specifically to a plurality of circulating  
 CC descendant endogenous retrovirus ancestor proteins), a method of  
 CC preparing an ancestral endogenous retroviral amino acid sequence, a  
 CC method for inducing an immune response to a donor virus in a transplant  
 CC recipient or a potential transplant recipient, a method of making a  
 CC vaccine, a method for detecting infection with an endogenous retrovirus  
 CC and a method for performing xenotransplantation in a subject. The  
 CC ancestral viral nucleic acid sequence is of Porcine Endogenous Retrovirus  
 CC (PERV) subtype A, B or C and has at least 70% identity with sequence of  
 CC any of the 18 nucleotide sequences fully defined in the specification.  
 CC The sequence may also be optimized for expression in a human host. The  
 CC nucleic acid sequence or its fragment is useful in the prophylaxis of  
 CC viral infection in transplantation that is heightened by the presence of  
 CC factors commonly associated with viral activation, e.g. immune  
 CC suppression, graft versus host disease, graft rejection, viral co-  
 CC infection, and cytotoxic therapies. The present sequence is an ancestral  
 CC PERV env protein.

XX Sequence 678 AA;

# Alignment Scores:

Pred. No.: 7,06e-282 Length: 678  
 Score: 3318.00 Matches: 634  
 Percent Similarity: 93.9% Conservative: 1  
 Best Local Similarity: 93.8% Mismatches: 3  
 Query Match: 95.7% Indels: 38  
 DB: 9 Gaps: 7

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x ADV28034 (1-678)

QY 1 ATGCATCCACGTTAAACCGGGCCACCTCCGATTCGGGTGGAAACCCGAAAGACTG 60  
 DB 1 MethisProThrLeuSerArgHisLeuProIleArgGlyGlyLeuProLysArgLeu 20  
 QY 61 AAAATCCCTTAAAGCTCGCTCCATCGCGTGGTTTCCTTACTCTCAATAACCTCTCAG 120  
 DB 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
 QY 121 ACTAATGGTATGCGCATAGGACAGCCTGAACCTCCCATAAACCTTATCTCTCACCTGG 180  
 DB 41 ThrAenGlyMetArgIleGlyAspSerLeuAenSerHisLysProLeuSerLeuThrTrp 60  
 QY 181 TTAATTACTAGTCCGGCAGAGATTAAATATCAACACTCAAGGGAGCTCCTTTA 240  
 DB 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAenAenThrGlnGlyAlaProLeu 80  
 QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTCAGATCAGTTATTCCTAGTCTG----- 294  
 DB 81 GlyThrTrpTrpProAspLeuTrValCysLeuArgSerValIleProSerLeuAenAsp 100  
 QY 295 ---ACCTCACCCCCAGATATCTCCATGCTCACGGATTTTATGTTGCCCCAGGACCACCA 351  
 DB 101 GlnThrSerProProAspIleLeuHisAlaHisGlyPheTrValCysProGlyProPro 120  
 QY 352 AATAATGAAACATATGCGGAAATCCCGAGAGATTCTTTTGTAAACAATGGAAGTGTGA 411  
 DB 121 AsnAenGlyLysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysVal 140  
 QY 412 ACCTCTAATGATGATATTGGAAATGGCCAACTCTCAGCAGGATAGGGTAAAGTTTCT 471  
 DB 141 ThrSerAenAspGlyTrpTrpLysTrpProThrSerGlnGlnAspArgValSerPheSer 160  
 QY 472 TATGTCAACACCTATACCAGCTCTCGGACAATTAAATTAC-----CTGACC 516  
 DB 161 TyrValAenThrTyThrSerSerGlyGlnPheAenTyGlyHisGlyArgTrpLeuThr 180  
 QY 517 TGG-----ATTAGAAGTGAAGCCCAAGTCTCTCTCTCAGAC 555  
 DB 181 TrpGlnGlnArgValGlnLysAspIleArgThrGlySerProLysCysSerProSerAsp 200  
 QY 556 CTAGATTACCTAAAAATAAGTTTCTACTGAGAAAGGAAAAACAAGAAATATCTAAATGG 615

DB 201 LeuAspTyLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleLeuLysTrp 220  
 QY 616 GTAAATCGTATGCTTCGGGAAATCGTATATATATGAGGCTCGGTAAACACACAGGCTCC 675  
 DB 221 ValAsnGlyMetSerTrpGlyMetValTyTrpGlyGlySerGlyLysGlnProGlySer 240  
 QY 676 ATTCTAACTATTCCCTCAAAATAAAC---CAGCTGGAGCCTCCAATGGCTATAGGACCA 732  
 DB 241 IleLeuThrIleArgLeuLysIleAsnThrGlnLeuGluProProMetAlaIleGlyPro 260  
 QY 733 AATACGGTCTTTCAGCGGTCAAGACCCCAACCAAGGACCA----- 774  
 DB 261 AsnThrValLeuThrGlyGlnArgProProThrGlnGlyProProHisAsnLeuProVal 280  
 QY 775 -----GGACCATCTCT-----AACATAACTCTCGA 801  
 DB 281 ProGlnGlyProSerProAsnProAspIleThrGlnSerAspTyAsnIleThrSerGly 300  
 QY 802 TCAGACCCCACT-----GAGTCTAACAGCAGCACTAAAAATGGGGCAAAA 846  
 DB 301 SerAspProThrAenThrProArgAsnGluSerAenSerThrThrLysMetGlyAlaLys 320  
 QY 847 CTTTTTAGCCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAGGCTACC 906  
 DB 321 LeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAenSerThrThrProGluAlaThr 340  
 QY 907 TCTTCTTGCTGCTATGCTTAGCTTCGGGGCCACCTTACTATGAAGNATGGCTAGAGA 966  
 DB 341 SerSerCysTrpLeuCysLeuAlaLeuGlyProProTyTrpGluGlyMetAlaArg 360  
 QY 967 GGGAAATCAATGACAAAGAACATAGAGACCAATGCACATGGGATCCCAAAATAAG 1026  
 DB 361 GlyLysPheAenValThrLysGluHisArgaspGlnCysThrTrpGlySerGlnAsnLys 380  
 QY 1027 CTTACCCCTTACTAGGTTTCTGGAAAAGGCACTTCGCATAGGAAAGGTTTCCCCCATCCCAC 1086  
 DB 381 LeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProProSerHis 400  
 QY 1087 CAACACCTTTGTACCACTGAAGCCCTTAACTCAACCTCTGAGAGTCAATATCGTA 1146  
 DB 401 GlnHisLeuCysAenHisThrGluAlaPheAenGlnThrSerGluSerGlnTyLeuVal 420  
 QY 1147 CCTGTTATGACAGCTGGTGGGCATGTAATCTGGATTAACCCCTTGTTGTTTCCACCTTG 1206  
 DB 421 ProGlyTyAspAArgTrpTrpAlaCysAenThrGlyLeuThrProCysValSerThrLeu 440  
 QY 1207 GTTTTAAACCAACTAAAGATTTTTCATTATGTCCTCAAAATGTTTCCCGAGTGATTAC 1266  
 DB 441 ValPheAenGlnThrLysAspPheCysIleMetValGlnIleValProArgValTyTrp 460  
 QY 1267 TATCCGAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGAACAAGAGACA 1326  
 DB 461 TyrProGluLysAlaIleLeuAspGluTyAspTyArgAenHisArgGlnLysArgGlu 480  
 QY 1327 CCCATATCTCGACACTTGTCTGCTGCTCGGACTTGGAGTGGCAGCAGGTGTAGGAACA 1386  
 DB 481 ProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIleGlyValGlyThr 500  
 QY 1387 GGAACAGCTGCTCGGTTCACGGGACCCACAGCAGCTAGAAAACAGGACTTAGTAACCTACAT 1446  
 DB 501 GlyThrAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAenLeuHis 520  
 QY 1447 CGAATTGTAAACAGAGATCTCAAGCCCTAGAAAAATCTGTCAGTAACTCGAGGAATCC 1506  
 DB 521 ArgIleValThrGluAspLeuGlnAlaLeuGlySerValSerAenLeuGluGluSer 540  
 QY 1507 CTAACTCTTTATCTGAAGTAGTCTCTACAGAAATAGAAGGCTTAGATTTATTATTCTTA 1566  
 DB 541 LeuThrSerLeuSerGluValValLeuGlnAenArgArgGlyLeuAspLeuPheLeu 560  
 QY 1567 AAAGAAGGAGGATTATTGTGTAGCCCTTGAAAGGAGGAATGCTGTTTTTATGTGGATCATTTCA 1626

D	b		561	LysGluGlyGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAspHisSer	580
Q	y		1627	GGGGCCATCAGAGACTCCATAACAAGCTTAGAGAAAGTTGGAGAAGCGTCGAAGGAA	1686
D	b		581	GlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArgGlu	600
Q	y		1697	AAGGAACTACTCAAGGTGGTTTGAGGGATGGTTCAACAGGTCTCTTTGGTTGGCTACC	1746
D	b		601	LysGluThrTrhGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeuAlaThr	620
Q	y		1747	CTACTTCTCGCTTTAAACAGGACCCCTTAATAGTCTCTCCTCGTTACTCACAGTTGGGCCA	1806
D	b		621	LeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrValGlyPro	640
Q	y		1807	TGTATTATTAAAGTGAATTCCTTCATTAGAGAACCAATAAGTCAGTCCAGATCATG	1866
D	b		641	CysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMet	660
Q	y		1867	GTACTTAGACAAACAGTACCAAGCCGCTCTAGCAGGGAAGCTGGCCGC	1914
D	b		661	ValLeuArgGlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg	676
 RESULT 15 ABB82635					
ID			ABB82635 standard; protein; 653 AA.		
XX					
AC			ABB82635;		
XX					
DT			19-FEB-2003 (first entry)		
XX					
DE			PERV envelope protein.		
XX					
KW			Porcine endogenous retrovirus; PERV-A; virucide; infection; viral; PERV;		
KX			envelope protein.		
XX					
OS			Porcine endogenous retrovirus.		
XX					
FN			WO200286060-A2.		
XX					
PD			31-OCT-2002.		
XX					
PF			19-APR-2002; 2002WO-US012085.		
XX					
PR			20-APR-2001; 2001US-0285103P.		
XX					
PR			21-DEC-2001; 2001US-00029656.		
XX					
PA			(IMME-) IMMERGE BIOTHERAPEUTICS INC.		
XX					
PI			Patience C, Ericsson T, Oldmixon BA;		
XX					
DR			WPI; 2003-067636/06.		
XX					
DR			N-PSDB; ABV75053.		
XX					
PT			New porcine endogenous retrovirus (PERV)-A receptor polypeptides and		
XX			polynucleotides, useful for screening agents that block viral infection		
PT			(which are useful for inhibiting PERV infection) or detecting tissues		
XX			infected with PERV.		
PT					
XX					
PS			Disclosure; Page 54-56; 71pp; English.		
XX					
CC			The invention relates to Porcine endogenous retrovirus (PERV)-A receptor		
XX			polypeptides found on primate cells including human and baboon and		
CC			encoding polynucleotides. The polypeptides are useful in both in vitro or		
XX			in vivo screening assays and models for agents that block viral		
CC			infection. These are also useful for detecting tissues infected with		
XX			PERV. Anti-PERV antibodies are useful for inhibiting PERV infection in		
CC			tissue, particularly in tissues for transplantation. The present sequence		
XX			represents the PERV-A envelope protein		
SQ			Sequence 653 AA;		
Alignment Scores:		6.4e-267		Length: 653	
Pred. No.:					

```
|||||
Db      321  TrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAlaAraArgGlyLysPhe 340
QY      976  AATGTGACAAAGAACATAGAGACCAATGACATGGGATCCCAAAATAAGCTTACCCCTT 1035
Db      341  AsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGlnAsnLysLeuThrLeu 360
QY     1036  ACTGAGGTTTCTGGAAAAGGCACCTGCATAGGAAGGTTCCCCCATCCCAACACACCTT 1095
Db      361  ThrGluValSerGlyLysGlyThrCysIleGlyLysValProProSerHisGlnHisLeu 380
QY     1096  TGTAAACCACTGAAGCCCTTAAATCAAAACCTCTGAGAGTCAATATCTGGTACCTGGTTAT 1155
Db      381  CysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyrLeuValProGlyTyr 400
QY     1156  GACAGGTGGTGGGCATGTAATCTGGATTAAACCCCTGTGTTCACCTTGGTTTTTAAC 1215
Db      401  AspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeuValPheAsn 420
QY     1216  CAAACTAAGATTTTGGCATTATGGTCCAAATTGTTCCCGAGTGTATTACTATCCCGAA 1275
Db      421  GlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrTyrProGlu 440
QY     1276  AAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAGAGAGAACCCATATCT 1335
Db      441  LysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLysArgGluProIleSer 460
QY     1336  CTGACACTTGCTGTGATCTCGACTTGGAGTGGCAGAGGTGTAGGAACAGGAACAGCT 1395
Db      461  LeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyValGlyThrGlyThrAla 480
QY     1396  GCCCTGTCACGGGACCAACGACGCTAGAACAGGACTTAGTAACTACATCGAATTGTA 1455
Db      481  AlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHisArgIleVal 500
QY     1456  ACAGAAAGATCTCAAGCCCTAGAAAAATCTGTCACTAAGTGGAGGAATCCCTAACCTCC 1515
Db      501  ThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluGluSerLeuThrSer 520
QY     1516  TTATCTGAAGTAGTCCTCAGAAATAGAAGAGGTTAGATTATTATTCTAAAAAGAGA 1575
Db      521  LeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuPheLeuLysGluGly 540
QY     1576  GGATTTATGTGTAGCCCTTGAAGGNGAATGCTGTTTTTATGTGGATCATTCAGGGCCATC 1635
Db      541  GlyLeuCysValAlaLeuLysGluCysCysPheTyrValAspHisSerGlyAlaIle 560
QY     1636  AGAGACTCCATGAACAAGCTTAGAAGAGGTTGGAGAGCGTCCGAAGGGAAGGAAACT 1695
Db      561  ArgAspSerMetSerLysLeuArgGluArgLeuGluLysArgArgGluLysGluThr 580
QY     1696  ACTCAAGGGTGGTTTGAAGGATGGTTCAACAGGTCTCTTTGGTTGGCTACCCCTACTTCT 1755
Db      581  ThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeuAlaThrLeuLeuSer 600
QY     1756  GCTTTTAAAGGACCCCTTAATAGTCTCTCTCTGTTACTCAGCTTGGGCCCATGTATTAT 1815
Db      601  AlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrValGlyProCysIleIle 620
QY     1816  AACAGTTTAATTCCTTCATTAGAGAACGAATAAGTCAGTCCAGATCATGTGTACTTAGA 1875
Db      621  AsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMetValLeuArg 640
QY     1876  CAACAGTACCAAGCCCTCTAGCAGGGAAGCTGCCCGC 1914
Db      641  GlnGlnTyrGlnSerProSerSerArgGluAlaGlyArg 653
```

GenCore version 5.1.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 16:07:51 ; Search time 48.1266 Seconds  
(without alignments)  
3323.427 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_5620\_7533

Perfect score: 3468

Sequence: 1 ATGCATCCAGCTTAACCG.....CTAGCAGGAGCTGGCGCGC 1914

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abss/ABSSWEB spool/US10723552/runat 14022006 125148 13205/app\_query.fasta\_1  
-DB=Published Applications AA.Main -QPMT=fastan -SUFFIX=rapbm -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum2  
-TRANS-human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abss02p  
-USER=US10723552 @CNG 1.1 602 @runat 14022006 125148 13205 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG -DEV TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA.Main:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3420	98.6	638	3	US-09-851-859A-6
2	3407	98.2	638	4	US-10-441-949-32
3	3407	98.2	638	4	US-10-441-949-34
4	3407	98.2	638	4	US-10-441-949-40
5	3407	98.2	638	4	US-10-441-949-42
6	3407	98.2	638	4	US-10-441-949-45
7	3318	95.7	678	4	US-10-441-949-36
8	3318	95.7	678	4	US-10-441-949-38
9	3147.5	90.8	653	4	US-10-029-656-2
10	3135.5	90.4	653	4	US-10-029-656-4
11	2879.5	83.0	661	4	US-10-441-949-43

12	2874.5	82.9	660	4	US-10-441-949-16
13	2871.5	82.8	660	4	US-10-441-949-8
14	2866.5	82.7	660	3	US-09-851-859A-5
15	2851	82.2	678	4	US-10-441-949-12
16	2849.5	82.2	660	4	US-10-441-949-10
17	2849.5	82.2	660	4	US-10-441-949-18
18	2835	81.1	678	4	US-10-441-949-14
19	2465	70.8	678	4	US-10-441-949-26
20	2455	70.8	678	4	US-10-441-949-24
21	2432	70.1	657	4	US-10-441-949-22
22	2432	70.1	657	4	US-10-441-949-30
23	2432	70.1	658	4	US-10-441-949-44
24	2422	69.8	657	3	US-09-851-859A-4
25	2422	69.8	657	4	US-10-441-949-20
26	2422	69.8	657	4	US-10-441-949-28
27	2353	67.8	656	3	US-09-851-859A-3
28	1520	43.8	676	4	US-10-677-558-3
29	1489.5	42.9	667	3	US-09-070-630-13
30	1489.5	42.9	667	5	US-10-732-438-13
31	1347	38.8	672	3	US-09-954-983-4
32	1346	38.8	675	3	US-09-954-983-6
33	1346	38.8	675	3	US-09-954-983-8
34	1345	38.8	672	3	US-09-954-983-5
35	1345	38.8	676	3	US-09-954-983-11
36	1343	38.7	675	3	US-09-954-983-7
37	1232.5	35.5	610	4	US-10-670-695-30
38	654	18.9	438	3	US-09-833-245-137
39	652	18.8	438	3	US-09-833-245-138
40	643	18.5	427	4	US-10-108-260A-4490
41	636	18.3	611	4	US-10-670-695-32
42	414.5	12.0	538	3	US-09-902-535-2
43	414.5	12.0	538	4	US-10-114-893-135
44	414.5	12.0	538	4	US-10-016-249-4
45	413.5	11.9	538	4	US-10-133-036-2

#### ALIGNMENTS

##### RESULT 1

US-09-851-859A-6  
; Sequence 6, Application US/09851859A  
; Patent No. US20020065407A1  
; GENERAL INFORMATION:  
; APPLICANT: Banerjee, Papia T.  
; APPLICANT: Patience, Clive  
; APPLICANT: Anderson, Goran K.  
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of  
; TITLE OF INVENTION: Use  
; FILE REFERENCE: 61750-321  
; CURRENT APPLICATION NUMBER: US/09/851.859A  
; CURRENT FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/097,015  
; PRIOR FILING DATE: 1998-08-18  
; PRIOR APPLICATION NUMBER: US 09/376781  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PERV-C  
; OTHER INFORMATION: polypeptide sequence taken from GenBank Accession  
; OTHER INFORMATION: No. US20020065407A1 AF038600 for comparison.  
US-09-851-859A-6

Alignment Scores:  
Pred. No.: 1.97e-312 Length: 638  
Score: 3420.00 Matches: 637  
Percent Similarity: 100.0% Conservatives: 1  
Best Local Similarity: 99.8% Mismatches: 0  
Query Match: 98.6% Indels: 0

DB: 3 Gaps: 0  
US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-09-851-859A-6 (1-638)  
QY 1 ATGCATCCACGTTAAACCGCGCCACCTCCGATTCGGGFTGGAAACCCGAAAGACTG 60  
Db 1 MethisProThrLeuAsnArgHisLeuProIleArgGlyGlyLeuProLysArgLeu 20  
QY 61 AAAATCCCTTAAGCTTCGCTCCATCCGCTGGTTCCTTACTCTGTCAATAACCTCTCAG 120  
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSergIn 40  
QY 121 ACTAATGTATGCGCATAGGACAGCCTGAACTCCCATAAACCTTATCTCTCACCTGG 180  
Db 41 ThrIleGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 181 TTAATTACTGCTCCGGCACAGGTATTAAATATCAACAACACTCAAGGGAGGCTCTTTA 240  
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80  
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTCAGATCAGTATTTCCTAGTCTGACCTCA 300  
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
QY 301 CCCCAGATATCCTCATGCTCACGGATTTTATGTTTCCCGCAGGACCAACCAATAATGGA 360  
Db 101 PropProAspIleLeuHisGlyPheIleGlyPheIleValCysProGlyProProAsnAsnGly 120  
QY 361 AAACATTCGGGAAATCCCGAGATTTCTTTTGTAAACAATGGAACCTGTAACTCTAAT 420  
Db 121 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 140  
QY 421 GATGATATTGGAATGGCCAACTCTCAGCGAGTAGGTAAGTTTCTTATGTCAAC 480  
Db 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
QY 481 ACCTATACCGCTCTGGCAATTTAATTACCTGACCTGATTTAGAACTGGAGCCCAAG 540  
Db 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180  
QY 541 TGCTCTCCTCAGACTAGATTACCTAAAAATAAGTTTCTACTGAGAAAGGAAACAAGAA 600  
Db 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
QY 601 AATATCTAAATGGGTAAATGGTATCTTCGGGAATGGTATATTATGGAGGCTCGGCT 660  
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyLysGly 220  
QY 661 AAACAACAGGCTCCATCTAATCTAATTCGCTCAAAATAAACCAGCTGGAGCTCCAATG 720  
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProPromet 240  
QY 721 GCTATAGGACCAAAATACGGTCTTGACGGGTCAAAGACCCCAACCAAGGACCCAGGACCA 780  
Db 241 AlalieGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
QY 781 TCCTCTAATCAATCTTGATCAGACCCCACTGAGTCTTAACAGACAGCTAAATATGGG 840  
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerSerSerThrThrLysMetGly 280  
QY 841 GCAAAATTTTATAGCTCATCCAGGAGCTTTTCAAGCTCTTAAGCTTCAAGCTCCAGACTCCAG 900  
Db 281 AlalysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
QY 901 GCTACTCTTCTTGTGTAGCTTACTTGGGCCCACTTATCGGGCCCACTTACTATGAAGGAATGGCT 960  
Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGlyGlyMetAla 320  
QY 961 AGAAGAGGAAATCAATGTGCAAAAGAACATAGAGACCAATGACATGGGATGCCAA 1020  
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
QY 1021 AATAAGCTTACCTTACTGAGGTTTCTGGAAGGACCTGTCATAGGAAAGGTTCCCCCA 1080

Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 1081 TCCACACAAACCTTTGTAAACCACTGAAGCCCTTTAATCAAACTCTGAGAGCAATAT 1140  
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
QY 1141 CTGCTACTGCTGATGACAGGTGGTGGCATGTAATACTGATTAACCCCTTGTGTTTCC 1200  
Db 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 1201 ACCTTGGTTTTTAACCAACTAAAGATTTTTCGCAATTATGGTCCAAATTTGTTCCCGAGTG 1260  
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 1261 TATTACTATCCGGAAGCAATCCTTGATGAATATGACTACAGAAATCATCGCAAAAG 1320  
Db 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
QY 1321 AGAGAACCATATCTCTGACACTTGTGTGATGCTCGGACTTGGAGTGGCAGCAGGTGTA 1380  
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
QY 1381 GGAACAGGAAACAGCTGCCCTGTCACCGGACCCAGCAGCTAGAAACAGGACTTAGTAAC 1440  
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 1441 CTACATCGAATGTAAACAGAAAGATCTCAAGCCCTAGAAAAATCTGTCAAGTAACCTGAG 1500  
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 1501 GAATCCCTTAACCTCTTCTGAAGTGTAGTCTCAGAGTGTAGAGAGGCTTAGATTTATTA 1560  
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeu 520  
QY 1561 TTTCTAAAGAGGAGGATTTATGTAGTCCCTTGAAGGAGGAATGCTGTTTTTATGTGGAT 1620  
Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGlyCysPheTyrValAsp 540  
QY 1621 CATTCAGGGGCCATCAGAGACTCCATGAAACAAGCTTAGAGAAAGGTTGGAGACGCTCGA 1680  
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560  
QY 1681 AGGGAAGAGGAAACTACTCAAGGGTGGTTTCAGGAGTGGTTCAACAGGTCTCTTTGGTTG 1740  
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerLeuTrpLeu 580  
QY 1741 GCTACCCCTACTTCTGCTTTAACAGGACCCCTTAATAGTCTCTCTCTCTTACTCACAGTT 1800  
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
QY 1801 GGGCCATGTATTATTAAACAGTTAATTCGCTTTCATTAGAGAACGATTAAGTCAGTCCAG 1860  
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgGlySerAlaValGln 620  
QY 1861 ATCATGTACTTAGACACAGTACCAGGCCCTCTAGCAGGAGCTGGCGCG 1914  
Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg 638  
RESULT 2  
US-10-441-949-32  
; Sequence 32, Application US/10441949  
; Publication No. US2004011684A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodrigo, Allen  
; APPLICANT: Ross, Howard A.  
; APPLICANT: Mullins, James I.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES  
; FILE REFERENCE: 08987-012001  
; CURRENT APPLICATION NUMBER: US/10/441,949  
; CURRENT FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: PCT/US01/05288  
; PRIOR FILING DATE: 2001-02-16



```
QY 1861 ATCATGTACTTAGAACACAGTACCAAGCCCTCTAGCAGGAAGCTGGCCGC 1914
Db 621 IImetValLeuArgGlnGlnTrpGlnSerProSerArgGluAlaGlyArg 638

RESULT 3
US-10-441-949-34
; Sequence 34, Application US/10441949
; Publication No: US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodriago, Allen
; APPLICANT: Mullins, James I.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-34

Alignment Scores:
Pred. No.: 3,29e-311 Length: 638
Score: 3407.00 Matches: 635
Percent Similarity: 99.7% Conservative: 1
Best Local Similarity: 99.5% Mismatches: 2
Query Match: 98.2% Indels: 0
DB: 4 Gaps: 0

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-10-441-949-34 (1-638)
QY 1 ATGCATCCACGGTTAAACCGCGCCACCTCCCGATTCGGGTGGAAAGCCGAAAGACTG 60
Db 1 MethisProThrLeuSerArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20
QY 61 AAAATCCCTTAAGTCCTCCATCGCGTGTTCCTTACTCTGTCATTAACCTCTCAG 120
Db 21 LysileProLeuSerPheAlaSerIleAlaIleArgPheLeuThrLeuSerIleThrSerGln 40
QY 121 ACTAATGTATGCCATAGGAGACAGCTGAACTCCCATAAACCTTATCTCTCACCTGG 180
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTAATGACTCCGGCAGCAGGTATTAATATCAACACACTCAAGGGGAGGCTCTTTTA 240
Db 61 LeuileThrAspSerGlyThrGlyLeuAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80
QY 241 GGAACCTGGTGGCTGATCTATAGCTTTGGCTTCAGATCAGTATTCTAGTCTGACCTCA 300
Db 81 GlyThrTrpTrpProAspLeuTyValCysLeuArgSerValIleProSerLeuThrSer 100
QY 301 CCCCAGATATCTCCATGCTCACGATTTTATGTTTGGCCAGGACACACCAAAATAATGGA 360
Db 101 ProProAspIleLeuHisAlaHisGlyPheTyValCysProGlyProProAsnAsnGly 120
QY 361 AAACATTGCGGAAATCCAGAGATTTCTTTTGTAAACAATCGAACTGTGTAACCTCTAAT 420
Db 121 LysHisCysGlyAsnProArgAspPheCysLysGlnTrpAsnCysValThrSerAsn 140
QY 421 GATGATATTGGAATGCGCAACCTCTCAGCAGGATAGGTTAGTTTCTTATGTCAAC 480
Db 141 AspGlyTyTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyValAsn 160
QY 481 ACCTATACCAGCTCTGGCAATTTAAATTACCTGACCTGGATTAGAACTGGAAAGCCCAAG 540
```

```
Db 161 ThrTyThrSerSerGlyGlnPheAsnTyLeuThrTrpIleArgThrGlySerProLys 180
QY 541 TGCCTCTCTTCAGACCTAGATTACCTTAAATAAATAGTTTCTACGTGAGAAAGGAAACAGAA 600
Db 181 CysSerProSerAspLeuAspTyLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200
QY 601 AATATCCTAAATGGGTAAATGGTATGTCCTTGGGAATGGTATATATATATGAGGGCTCCGGT 660
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyTrpGlyGlySerGly 220
QY 661 AAACAACACAGGCTCCATCTTAACCTATTTCGGCTCAAAATAAAACACAGCTGGAGCTCCCAATG 720
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240
QY 721 GCTATAGGACAAATACGGTCTTGACGGGTCAAGACCCCAACCAAGACACGAGACCA 780
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProThrGlnGlyProGlyPro 260
QY 781 TCCTCTAACATACTCTCGATCAGACCCCACTGAGTCTAACAGCAGCAGCTTAAATGGG 840
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280
QY 841 GCAAAACCTTTTTCAGCTCATCCAGGGAGCTTTTCAAGCTCTTAACTCCAGACTCCAGAG 900
Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300
QY 901 GCTACCTCTTCTTGGCTATGCTTAGCTTCGGGCCCACTTACTATAGAGGAATGGCT 960
Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyTrpGlyGlyMetAla 320
QY 961 AGAAGAGGGAAATCAATGTGACAAAGAACATAGACACCAATGCACATGGGATCCCAA 1020
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340
QY 1021 AATAAGCTTACCTTACTGAGGTTTCTGGAAAGGACCTGCAATAGGAAAGGTTTCCCCCA 1080
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360
QY 1081 TCCCAACAACACCTTTGTAAACCACTGAAAGCTTTTAACTCAACCTCTGAGAGTCAATAT 1140
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTy 380
QY 1141 CTGTACTCTGTTATGACAGTGTGGGCATGTAATACTGATTAACCCCTTGTGTTTCC 1200
Db 381 LeuValProGlyTyAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400
QY 1201 ACCTTGTGTTTTTAAACCAAACTAAAAGATTTTTCATTTATGGTCCAAATTTGTTCCCGAGTG 1260
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420
QY 1261 TATTACTATCCGAAAAAGCAATCCTTGATGATATGACTACAGAAATCATCGACAAAG 1320
Db 421 TyTrpTyTrpProGluLysAlaIleLeuAspGluTyAspTyArgAsnHisArgGlnLys 440
QY 1321 AGAAGACCAATATCTCTGACACTTGCCTGATGCTCGACTTGGAGTGGCAGCAGGTGTA 1380
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIleGlyVal 460
QY 1381 GGAACAGGAACAGCTGGCTGGTCCAGGGACCAACAGCAGCTAGAAAACAGGACTTAGTAAC 1440
Db 461 GlyThrGlyThrAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
QY 1441 CTACATCGAATTGACAGAGATCTCCAAGCCCTAGAAAAATCTGTCAGTAACTGGAG 1500
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGlyLysSerValSerAsnLeuGlu 500
QY 1501 GAATCCCTTAACCTCTTATCTGAAGTAGTCTCAGAAATAGAAGAGGTAGATTATTA 1560
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520
QY 1561 TTTCTAAAGAAGAGGAGGATTATGTGTAGCCCTTGAAGGAGGAATGCTGTTTTTATGTGGAT 1620
```



Db 521 PheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysPheTyrValAsp 540  
Qy 1621 CATTACGGGGCCATCAGAGACTCCATGAACAAGCTTACAGAAAAGTTGGAGAAGCGTCGA 1680  
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560  
Qy 1681 AGGGAAGGAACACTACTCAAGGTGGTTGAGGATGGTTCAACAGGTCTCTTTGGTTG 1740  
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580  
Qy 1741 GCTACCCCTACTTCTCGCTTTTAAACAGGACCTTAATAGTCCCTCTCTTACTCACAGTT 1800  
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrVal 600  
Qy 1801 GGGCCATGTATTATTAAACAAGTTAATTCCTTTCATTAGAGAACGAATAAGTGCAGTCCAG 1860  
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
Qy 1861 ATCATGGTACTTAGACAACAGTACCAGGACCGCTCTAGCAGGGAAGCTGGCCGC 1914  
Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg 638

RESULT 4

US-10-441-949-40  
; Sequence 40, Application US/10441949  
; Publication No. US20040116684A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodriago, Allen  
; APPLICANT: Ross, Howard A.  
; APPLICANT: Mullins, James I.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES  
; FILE OF INVENTION: 08987-012001  
; CURRENT APPLICATION NUMBER: US/10/441,949  
; PRIOR FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: PCT/US01/05288  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/183,659  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 40  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially generated peptide  
US-10-441-949-40

Alignment Scores:  
Pred. No.: 3,29e-311 Length: 638  
Score: 3407.00 Matches: 635  
Percent Similarity: 99.7% Conservative: 1  
Best Local Similarity: 99.5% Mismatches: 2  
Query Match: 98.2% Indels: 0  
DB: 4 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-10-441-949-40 (1-638)

Qy 1 ATGCATCCACGTTAAACCGCGCCACCTCCCGATTCGGGTGGAAGCCGAAAAGACTG 60  
Db 1 MetHisProThrLeuSerArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20  
Qy 61 AAAATCCCTTAAGCTTCGCTCCATCGCTGGTGTCTTACTCTGTCAATAACCTCTCAG 120  
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
Qy 121 ACTAATGGTATGGCATAGGAGACAGCTGAACTCCCATAAACCTTATCTCTCACCTGG 180  
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
Qy 181 TTAATTAAGTCCCGCACAGGTATTAAATATCAACAAGTCAAGGGAGGCTCCCTTTA 240  
Db 61 LeuIleThrAspSerGlyThrGlyIleAsnIleAsnThrGlnGlyGluAlaProLeu 80

Qy 241 GGAACCTGGTGGCCTGATCTATACGTTTGCCTCAGATCAGTTATTCTTAGTCTGACCTCA 300  
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
Qy 301 CCCCAGATATCTCCATGCTCAGGATTTTATGTTGCCAGGACCAACCAATAATGGA 360  
Db 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120  
Qy 361 AAACATTGCGGAATCCACAGAGATTTCTTTTGTAAACAATGAAGTGTGTAACTCTAAT 420  
Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140  
Qy 421 GATGGATATTGGAATGGCAACCTCTCAGCAGATAGGGTAAGTTTCTTATGTCAAC 480  
Db 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
Qy 481 ACCTATACAGCTCTGGACAAATTAATTAATCTGACCTGGATTAGAACTGGAAGCCCAAG 540  
Db 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180  
Qy 541 TGCTCTCTCTCAGACCTAGATTACCTAAAATAAGTTTCACTGAGAAGGAACAAGAA 600  
Db 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
Qy 601 AATATCTTAAATGGGTAAATGGTATGCTCTTGGGAATGGTATATATTATGAGGCTCGGGT 660  
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGly 220  
Qy 661 AAACAACAGGCTCCATTTCAATTTTCGCTCAAAATAAACACAGCTGGAGCTTCCAATG 720  
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
Qy 721 GCTATAGGACCAATACGCTCTGACGGGTCAAGACCCCCCAAGCCAGGACGAGGACCA 780  
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
Qy 781 TCCTCTAACATAAATCTCTGGATCAGACCCCACTGAGTCTAACACAGCAGCTTAAATGGG 840  
Db 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
Qy 841 GCAAAACCTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCAGACTCCAGAG 900  
Db 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
Qy 901 GCTACCTCTTCTTGGTCTATGCTTAGCTTCGGGCCACCTTACTACTATGAGGAATGGCT 960  
Db 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGlyGlyMetAla 320  
Qy 961 AGAAGAGGGAATTTCAATGTGACAAAAGAACATAGAGACCAATGCACATGGGATCCCAA 1020  
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340  
Qy 1021 AATAAGCTTACCTTACTAGAGTTTCTGGAAGGACACCTGCATAGAAAGTTTCCCCCA 1080  
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
Qy 1081 TCCCACCAACACCTTTGTAAACCACTGAAGCTTTAATCAAACTCTGAGAGTCAATAT 1140  
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
Qy 1141 CTGGTACCTGGTTATGACAGGTGCTGGCATGTAACTAGTAACTTAAACCTCTGGTTTCC 1200  
Db 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
Qy 1201 ACCTTGGTTTTTAAACCAACTAAAGATTTTTCATTATGGTCCAAATTTGTTCCCGAGTG 1260  
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
Qy 1261 TATTACTATCCCCGAAAAGCAATCCTTGATGATATATGACTACAGAAATCATCGACAAAAG 1320  
Db 421 TyrTyrTyrProGlyLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440

QY 1321 AGAGAACCCATATCTCTGACACTTGCTGATGCTCGGACTGGAGTGGCAGCAGGTGTA 1380  
Db |||||||  
QY 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
Db |||||||  
QY 1381 GGAACAGGAACAGCTCGCCCTGCTACGCGGACACACAGCAGCTAGAAACAGGACTTAGTAAC 1440  
Db |||||||  
QY 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
Db |||||||  
QY 1441 CTACATCGAATTGTAAACAGAGATCTCAAGCCCTAGAAAAATCTGTCAAGTAACCTGGAG 1500  
Db |||||||  
QY 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
Db |||||||  
QY 1501 GAATCCCTAACTCTCTGTAAGTCTGCTACAGGAGTGTACAGAGGCTGTAGATTATTATTA 1560  
Db |||||||  
QY 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
Db |||||||  
QY 1561 TTTCTAAAGAGGAGGATTTATGTAGCTTGTGAAGGAGGAATGCTGTTTTTGTGTGAT 1620  
Db |||||||  
QY 521 PheLeuLysGluGlyLeuCysValAlaLeuLysGluGluCysCysPheTyrValAsp 540  
Db |||||||  
QY 1621 CATTGAGGGGCATCAGAGCTCCATGAACAGCTTAGAGAAAGTTGGAGAGGCTGCA 1680  
Db |||||||  
QY 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560  
Db |||||||  
QY 1681 AGGGAAGGAAACTACTCAAGGTGTGTGAGGATGTTCAACAGGTCTCTTTGGTTG 1740  
Db |||||||  
QY 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580  
Db |||||||  
QY 1741 GCTACCTACTTCTGCTTTAAACAGGACCTTAATAGTCCCTCTCTCTACTCACAGTT 1800  
Db |||||||  
QY 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
Db |||||||  
QY 1801 GGGCCATGTTATTAAACAAGTTAATGCTTTCATTAGAGAACGAATAAGTGCAGTCCAG 1860  
Db |||||||  
QY 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
Db |||||||  
QY 1861 ATCATGTACTTAGACACAGTACCAGAGCCGCTAGCAGGAGCTGGCGC 1914  
Db |||||||  
QY 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg 638

RESULT 5

; Sequence 42, Application US/10441949  
; Publication No. US20040116684A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodrigo, Allen  
; APPLICANT: Ross, Howard A.  
; APPLICANT: Mullins, James I.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES  
; FILE REFERENCE: 08987-012001  
; CURRENT APPLICATION NUMBER: US/10/441,949  
; CURRENT FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: PCT/US01/05288  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/183,659  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 42  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially generated peptide  
US-10-441-949-42

Alignment Scores:  
Pred. No.: 3,29e-311 Length: 638  
Score: 3407.00 Matches: 635  
Percent Similarity: 99.7% Conservative: 1  
Best Local Similarity: 99.5% Mismatches: 2  
Query Match: 98.2% Indels: 0  
DB: 4 Gaps: 0

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-10-441-949-42 (1-638)  
QY 1 ATGCATCCAGCTTAAACCGCGCCACCTCCCGATCGGGTGGAAAGCCGAAAGACTG 60  
Db |||||||  
QY 1 MethIspProThrLeuSerArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20  
Db |||||||  
QY 61 AAAATCCCTTAAGCTTCGCTCCATCGCGTGTCTCTTACTCTGTCAATAACCTCTCAG 120  
Db |||||||  
QY 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
Db |||||||  
QY 121 ACTAATGTATGCTCCATAGGAGACAGCTGAACCTCCATAAACCTTATCTCTCACCTGG 180  
Db |||||||  
QY 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
Db |||||||  
QY 181 TTAATTACTGACTCCGCGCACAGGTATTAAATATCAACACTCAAGGGAGGCTCTTTA 240  
Db |||||||  
QY 61 LeuIleThrAspSerGlyThrGlyLeuAsnAsnThrGlnGlyGluAlaProLeu 80  
Db |||||||  
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGCCTCAGATCAGTTATTCTCTAGTCTGACCTCA 300  
Db |||||||  
QY 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
Db |||||||  
QY 301 CCCCACGATATCTCCCATGCTCACGGATTTTATGTTGCCAGGACACCAACAAATAATGGA 360  
Db |||||||  
QY 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120  
Db |||||||  
QY 361 AAACATTCGCGAAATCCAGAGATTCTTTTGTAAACAATGGAACGTGTGTAACCTCTAAT 420  
Db |||||||  
QY 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140  
Db |||||||  
QY 421 GATGATATTGGAATGGCAACCTCTCAGCAGATAGGTAGGTAAGTTTTCTTATGTCAAC 480  
Db |||||||  
QY 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
Db |||||||  
QY 481 ACCTATACCACTCTGGACAAATTTAATVACCTGAGCTGGATTAGAACTGGGAAGCCCCAAG 540  
Db |||||||  
QY 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180  
Db |||||||  
QY 541 TGCTCTCTTCAGACCTAGATTACTTAAATAAGTTTCTACTGAGAAAGGAAAAACAAGAA 600  
Db |||||||  
QY 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
Db |||||||  
QY 601 AATATCTTAAATCGGTAAATGTTATCTCTTGGGGAATGGTATATTATGGAGGCTCGAGT 660  
Db |||||||  
QY 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyLysSerGly 220  
Db |||||||  
QY 661 AAACAACCAAGCTCCATTCTAACTATTCCGCTCAAAAATAAACCAAGCTGGAGCTCCAATG 720  
Db |||||||  
QY 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
Db |||||||  
QY 721 GCTATAGACCAATATACGTTCTTGAAGGTCAAGAGACCCCAACCCCAAGGACCAGACCA 780  
Db |||||||  
QY 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyPro 260  
Db |||||||  
QY 781 TCCTCTAACATACTTCTGGATCAGCCCCACTGCTACAGCACCACCTAAATAATGGG 840  
Db |||||||  
QY 261 SerSerAsnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280  
Db |||||||  
QY 841 GCAAAACTTTTATAGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCAGACTCCAGAG 900  
Db |||||||  
QY 281 AlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300  
Db |||||||  
QY 901 GCTACCTCTTCTTGTGTGCTATGCTTAGCTTCGGGCCCACTTACTTACTAAGAGGAATGGCT 960  
Db |||||||  
QY 301 AlaThrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGlyGlyMetAla 320  
Db |||||||  
QY 961 AGAAGAGGGAATTCATGTGCAAAAAGAAACATAGAGACCAATGCACATGGGATCCCAA 1020  
Db |||||||  
QY 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpLysSerGln 340  
Db |||||||  
QY 1021 AATAAGCTTACCTTACTGAGGTTTCTGGAAGAGGCACCTCATAGGAAGAGGTTCCTCCCA 1080  
Db |||||||

Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360  
QY 1081 TCCCAACCAACCTTTGTAACCACTGAAGCTTTAAATCAAACTCTGAGAGTCAATAT 1140  
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380  
QY 1141 CTGGTACCTGGTTATGACAGGTGGTGGCATGTAACTACTGATTAACCCCTTGTTGCC 1200  
Db 381 LeuValProGlyTyrAspArgTrpAlaCysAsnThrGlyLeuThrProCysValSer 400  
QY 1201 ACCTTGGTTTAAACCAAACTAAAGATTTTGCTATTATGGTCCAAATTTGTTCCCGAGTG 1260  
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420  
QY 1261 TATTACTATCCGAAAAAGCAATCTTCGATGAATGACTACAGAAATCATCGACAAAG 1320  
Db 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440  
QY 1321 AGAGAACCATATCTGCACACTTCTGTGTGATGCTCGGACTTGGAGTGGCAGCGTGTA 1380  
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460  
QY 1381 GGAACAGGAACAGCTGCCCTCGTCACGGGACACAGCAGCTAGAAAACAGGACTTAGTAAC 1440  
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480  
QY 1441 CTACATCGAATGTGAACAGAGATCTCCAAGCCCTAGAAAAATCTGTCAAGTAACCTGGAG 1500  
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500  
QY 1501 GAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGAATAGAAGGGTTAGATTATTA 1560  
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520  
QY 1561 TTTCTTAAAGAGGAGGATTATGTAGCTTGAAGGAGGAATCTCTTTTATGTGAT 1620  
Db 521 PheLeuLysGluGlyLysCysValAlaLeuLysGluLysCysPheTyrValAsp 540  
QY 1621 CATTGAGGGCCATCAGAGACTCCATGAACAAAGCTTAGAGAAAGTTGGAGAAGCTCGA 1680  
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560  
QY 1681 AGGAAAGGAAACTACTCAAGGTGGTTTGAAGGATGGTTCAACAGGTCTCTTTGGTTG 1740  
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTyrLeu 580  
QY 1741 GCTACCTTACTTCTGCTTAAACAGGACCTTAAATAGTCCCTCTCTTACTCACAGTT 1800  
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600  
QY 1801 GGGCCATGTATTATTAACAAGTTAAATGGCTTTCATTAGAGAACGAATAAGTGCAGTCCAG 1860  
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620  
QY 1861 ATCATGTACTTAGACACAGTACCAAGCCCGTCTAGCAGGGAAGCTGGCCG 1914  
Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg 638

## RESULT 6

US-10-441-949-45  
; Sequence 45, Application US/10441949  
; Publication No. US20040116684A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodrigo, Allen  
; APPLICANT: Ross, Howard A.  
; APPLICANT: Mullins, James I.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES  
; FILE REFERENCE: 08987-012001  
; CURRENT APPLICATION NUMBER: US/10/441,949  
; CURRENT FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: PCT/US01/05288  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/183,659

; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45  
; LENGTH: 638  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially generated peptide  
US-10-441-949-45

Alignment Scores:  
Pred. No.: 3,29e-311 Length: 638  
Score: 3407.00 Matches: 635  
Percent Similarity: 99.7% Conservative: 1  
Best Local Similarity: 99.5% Mismatches: 2  
Query Match: 98.2% Indels: 0  
DB: 4 Gaps: 0

US-10-723-552-3\_copy\_5620\_7533 (1-1914) x US-10-441-949-45 (1-638)

QY 1 ATGCATCCACGCTTAAACCGCGCCACCTCCCGATTGGGTGGAAAGCCGAAAGACTG 60  
Db 1 MethisProthrLeuSerArgHisLeuProIleargGlyLysProLysArgLeu 20  
QY 61 AAAATCCCTTAACTTCCCTCCATCGCTCCATCGGTGTTCTTACTCTGTCAATAACCTCTCAG 120  
Db 21 LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40  
QY 121 ACTAATGGTATGCCATAGGAGACAGCTGAACTCCCATAAACCTTATCTCTCACCTGG 180  
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60  
QY 181 TTAATTACTCAGCTCCGCGACAGGTATTAAATCAACAACACTCAAGGGAGGCTCCTTTA 240  
Db 61 LeullethrAspSerGlyThrGlyIleAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80  
QY 241 GGAACCTGGTGGCTGTATCTATACGTTTGGCTCAGATCAGTATTCTTAGTCTCACCTCA 300  
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuThrSer 100  
QY 301 CCCCAGATATCTCCATGCTCAGGATTTTATGTTTCCCGAGGACCAACCAATAATGGA 360  
Db 101 ProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProProAsnAsnGly 120  
QY 361 AAACATTCGGAATCCACAGAGATTTCTTTGTAAACAATGGAACCTGTGTAACCTCTAAT 420  
Db 121 LysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysValThrSerAsn 140  
QY 421 GATGGATATTGGAAATGCCCAACCTCTCAGCAGGATAGGGTAACTTTTCTTATGTCAC 480  
Db 141 AspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSerTyrValAsn 160  
QY 481 ACCTATACAGCTCTGGACAAATTAATCCTGACCTGGATAGAACTGGAAGCCCAAG 540  
Db 161 ThrTyrThrSerSerGlyGlnPheAsnTyrLeuThrTrpIleArgThrGlySerProLys 180  
QY 541 TGCTCTCTTCCAGACCTAGATTACCTAAAATAAGTTTCACTGAGNAAGGAACAAAGAA 600  
Db 181 CysSerProSerAspLeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlu 200  
QY 601 AATATCTCAAAATGGGTAAATGGTATGCTTTGGGAATGGTATATTATGAGGCTCGGGT 660  
Db 201 AsnIleLeuLysTrpValAsnGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGly 220  
QY 661 AAACAACAGGCTCCATTCTAACTATTTCGCTCAAAATAAACCAAGCTGGAGCTCCCAATG 720  
Db 221 LysGlnProGlySerIleLeuThrIleArgLeuLysIleAsnGlnLeuGluProProMet 240  
QY 721 GCTATAGGACCAATAGGCTCTTGACGGGTCAAGACCCCCCAAGCCCAAGGACGAGGACCA 780  
Db 241 AlaIleGlyProAsnThrValLeuThrGlyGlnArgProThrGlnGlyProGlyPro 260

```
QY 781 TCCTCTAAATAAATCTTCGGATCAGACCCCACTAGTCTAAACAGCACGACTAAATGGG 840
Db 261 SerSerAlnIleThrSerGlySerAspProThrGluSerAsnSerThrThrLysMetGly 280
QY 841 GCAAAATCTTTTAGCCTCATCCAGGAGCTTTTCAAGCTCTTAACTCCACGACTCCAGAG 900
Db 281 AlalysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGlu 300
QY 901 GCTACCTCTTCTGTGTGCTATGCTTAGCTTCGGGCCCACTTACTACTAAGAAGGATGGCT 960
Db 301 AlathrSerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAla 320
QY 961 AGAAGAGGAATCAATGTGACAAAGAACATAGAGACCAATGCACATGGGGATCCCAA 1020
Db 321 ArgArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGln 340
QY 1021 AATAAGCTTACCTTACTGAGGTTTCTCGAAAAGCCACTGCATAGGAAAGGTTCCCCCA 1080
Db 341 AsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProPro 360
QY 1081 TCCACCAACACCTTTGTAAACACACTGAAGCCTTTAATCAAACTCTGAGAGTCAATAT 1140
Db 361 SerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyr 380
QY 1141 CTGGTACCTGTTATGACAGTGGTGGCATCTAATCTGATTAAACCCCTGTGTTTCC 1200
Db 381 LeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSer 400
QY 1201 ACCTTGGTGTTTTAAACAACTAAAGATTTTTCGATTATGGTCCAAATTTTCCCGAGTG 1260
Db 401 ThrLeuValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgVal 420
QY 1261 TATTACTATCCGAAAAGCAATCTTGATGATATGACTACAGAAATCATCGCAAAAG 1320
Db 421 TyrTyrTyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLys 440
QY 1321 AGAGAACCATATCTCTCACACTGCTGTGATGCTCGGACTTGGAGTGCAGCAGGTGTA 1380
Db 441 ArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyVal 460
QY 1381 GAAACAGGAACAGCTGCCCTGGTCAAGGGACCAACAGCAGCTAGAAACAGGACTTAGTAAC 1440
Db 461 GlyThrGlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsn 480
QY 1441 CTACATCGAATGTAAACAGAGATCTCAAGCCCTAGAAAATCTGTGAGTAACTCGAG 1500
Db 481 LeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGlu 500
QY 1501 GAATCCCTAACTCTTATCTGAAGTCTCTACAGAAATAGAAGAGGTTAGATTATTA 1560
Db 501 GluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeu 520
QY 1561 TTTCTTAAAGAGGAGGATTTATGTAGCTTTGAAGGAGGAATGCTGTTTTTATGTGGAT 1620
Db 521 PheLeuLysGluGlyLeuCysValAlaLeuLysGluGluCysPheTyrValAsp 540
QY 1621 CATTCAAGGGCCATCAGAGATCTCATGAACAGCTTAGAGAAAGGTTGGAGAGCGTCGA 1680
Db 541 HisSerGlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArg 560
QY 1681 AGGGAAGGAAGAACTACTCAAGGTTGTTGAGGATGTTCAACAGGTCCTTTGGTTG 1740
Db 561 ArgGluLysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeu 580
QY 1741 GCTACCTTACTTTCTGTCTTAAACAGGACCTTAAATAGTCTCTCTCTCTCTCTCTCT 1800
Db 581 AlaThrLeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrVal 600
QY 1801 GGGCCATGATATTATTAACAAGTTAATGCTTTCATTAGAGAACGAATAGTGCAGTCCAG 1860
Db 601 GlyProCysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGln 620
QY 1861 ATCATGGTACTTTAGACAACAGTACCAGGACCGCTCTAGCAGGGAAGCTGGCCGC 1914
```

```
Db 621 IleMetValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg 638
RESULT 7
US-10-441-949-36
; Sequence 36, Application US/10441949
; Publication No. US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodrigo, Allen
; APPLICANT: Ross, Howard A.
; APPLICANT: Mullins, James I.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-36
Alignment Scores:
Pred. No.: 8,02e-303 Length: 678
Score: 3318.00 Matches: 634
Percent Similarity: 93.8% Conservative: 1
Best Local Similarity: 93.8% Mismatches: 3
Query Match: 95.7% Indels: 38
DB: 4 Gaps: 7
US-10-723-552-3_COPY_5620_7533 (1-1914) x US-10-441-949-36 (1-678)
QY 1 ATGCATCCAGCTTAAACCGCGGCACCTCCGATTCGGGTGGAAGCCGAAAGACTG 60
Db 1 MethisProThrLeuSerArgHisLeuProIleArgGlyLysProLysArgLeu 20
QY 61 AAAATCCCTTAAAGCTCGCTCCATCGCGGTTTCTTACTCTGTCTAATAACCTCTCAG 120
Db 21 LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40
QY 121 ACTAATGTGTATGCGATAGGACAGCGCTGAACCTCCATAAACCTTATCTCTCACCTGG 180
Db 41 ThrAsnGlyMetArgIleGlyAspSerLeuAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTAATGCTCGGCACAGGATTAATATCAACAACACTCAAGGGAGGCTCCTTTA 240
Db 61 LeuIleThrAspSerGlyThrGlyLeuAsnIleAsnAsnThrGlnGlyGluAlaProLeu 80
QY 241 GGAACCTTGGCTGCTGATCTATAGCTTTCCTCCATCAGATCAGTTATTCCTAGTCTG 294
Db 81 GlyThrTrpTrpProAspLeuTyrValCysLeuArgSerValIleProSerLeuAsnAsp 100
QY 295 ---ACCTCACCCCGCAGATATCTCATGCTCACCGATTATTTATGTTTGCACGACCA 351
Db 101 GlnThrSerProProAspIleLeuHisAlaHisGlyPheTyrValCysProGlyProPro 120
QY 352 AATAATGAAACATTCGCGAAATCCCGAGATTTCTTTGTAAACAATGAACGTGTGTA 411
Db 121 AsnAsnGlyLysHisCysGlyAsnProArgAspPhePheCysLysGlnTrpAsnCysVal 140
QY 412 ACCTCTAATGATGATATTGGAATGGCCAACTCTCAGCAGGATAGGTTAGTTTCT 471
Db 141 ThrSerAsnAspGlyTyrTrpLysTrpProThrSerGlnGlnAspArgValSerPheSer 160
QY 472 TATGTCAACCTATACCAGCTCTGGACAATTTAATTAC-----CTGACC 516
|||||
```

Db 161 TyrValAsnThrTyrThrSerSerGlyGlnPheAsnTyrGlyHisGlyArgTrpLeuThr 180  
Qy 517 TGG-----ATTAGAACTGGAAGCCCAAGTGTCTCTCTTCAGAC 555  
Db 181 TrpGlnGlnArgValGlnLysAspIleArgThrGlySerProLysCysSerProSerAsp 200  
Qy 556 CTAGATTACCTAAATAAGTTTCACTCAGAAAGGAAACAAGAAATATCTTAAATGG 615  
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleLeuLysTrp 220  
Qy 616 GTAATATGTTATGCTTGGGGGAATGTTATATTATGAGGCTCGGTAAACAACACCGGCTCC 675  
Db 221 ValAsnGlyMetSerTrpGlyMetValTyrTyrGlySerGlyLysGlnProGlySer 240  
Qy 676 ATTTAACTATTCGCTCAAAATAAAC---CAGCTGGAGCCTCAATGGCTTATAGGACCA 732  
Db 241 IleLeuThrIleArgLeuLysIleAsnThrGlnLeuGluProProMetAlaIleGlyPro 260  
Qy 733 AATAGGCTTGTGAGGGTCAAGAGCCCAACCCCAAGGACCA-----774  
Db 261 AsnThrValLeuThrGlyGlnArgProProThrGlnGlyProProHisAsnLeuProVal 280  
Qy 775 -----GGACCATCTCT-----AACATAACTTCTGGA 801  
Db 281 ProGlnGlyProSerProAsnProAspIleThrGlnSerAspTyrAsnIleThrSerGly 300  
Qy 802 TCAGACCCCACT-----GAGTCTAACAGCAGCACTAAATGGGGGCAAAA 846  
Db 301 SerAspProThrAsnThrProArgAsnGluSerAsnSerThrThrLysMetGlyAlaLys 320  
Qy 847 CTTTTAGCTCATCGAGGAGCTTTTCAAGCTCTTAACCTCCAGCACTCCAGAGGCTACC 906  
Db 321 LeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThr 340  
Qy 907 TCTTCTTGTGGCTATGCTTGGGCGCCCTTACTATGAGCAATGAGCAATGGCTAGAGA 966  
Db 341 SerSerCysTrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAlaArgArg 360  
Qy 967 GGGAAATTCATGTGACAAAGAAACATAGAGACCAATGCATGGGATCCCAAAATAAG 1026  
Db 361 GlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGlnAsnLys 380  
Qy 1027 CTTACCTTACTGAGGTTCTGGAAGGACGACCTGCATAGAAAGTTCCCCCACTCCAC 1086  
Db 381 LeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyLysValProProSerHis 400  
Qy 1087 CAACACCTTGTAAACACACTGAGCCTTTAATCAACCTCTGAGACTCAATATCTGTA 1146  
Db 401 GlnHisLeuCysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyrLeuVal 420  
Qy 1147 CCTGGTTATGACAGGTGGGCGCATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 1206  
Db 421 ProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeu 440  
Qy 1207 GTTTTAAACCAACTAAAGATTTTGTATGATGATGATGATGATGATGATGATGATGATG 1266  
Db 441 ValPheAsnGlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrTyr 460  
Qy 1267 TATCCCGAAGAAAGCAATCTTGATGATGATGATGATGATGATGATGATGATGATGATG 1326  
Db 461 TyrProGluLysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLysArgGlu 480  
Qy 1327 CCCATATCTCTGACACTTGTGTGATGCTCGGACTTGGAGTGGCAGCAGGTGTAGGAACA 1386  
Db 481 ProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValIleAlaGlyValGlyThr 500  
Qy 1387 GGAACAGCTGCCCTGGTACGGGACCAACAGCAGCTAGAAAACAGACTTACTACCTACAT 1446  
Db 501 GlyThrAlaAlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHis 520  
Qy 1447 CGAATTGTACAGAGATCTCCAGCCCTAGAAAAATCTGTAGTACCTGGAGGATCC 1506  
Db 521 ArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluGluSer 540

Qy 1507 CTAACCTCCTTATCTGAAGTAGTCTCTACAGAAATAGAGAGGTTAGATTATTATTCTTA 1566  
Db 541 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeu 560  
Qy 1567 AAGAACGAGGATTTATGTAGCTTGAAGAGGAAATGCTTTTATGTGGATCATTCATCA 1626  
Db 561 LysGluGlyGlyLeuCysValAlaLeuLysGluCysCysPheTyrValAspHisSer 580  
Qy 1627 GGGCCCATCAGAGACTCCATGAACAAGCTTAGAGAAAGGTTGGAGAGCGTCGAAGGAA 1686  
Db 581 GlyAlaIleArgAspSerMetAsnLysLeuArgGluArgLeuGluLysArgArgGlu 600  
Qy 1687 AAGAAACTACTCAAGGGTGTGTGAGGATGTTCAACAGGCTCTTTTGGTTGGCTACC 1746  
Db 601 LysGluThrThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeuAlaThr 620  
Qy 1747 CTACTTCTCTGTTTAAACAGGACCTTAATAGTCTCTCTCTGTTACTCAGAGTGGGCA 1806  
Db 621 LeuLeuSerAlaLeuThrGlyProLeuIleValLeuLeuLeuLeuThrValGlyPro 640  
Qy 1807 TGTATTATTAAAGTTAATTTGCCTTCATTAGAGAACGAATAAGTCAGTCCAGATCATG 1866  
Db 641 CysIleIleAsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMet 660  
Qy 1867 GTACTTAGACAACAGTACCAAGCCCTCTAGCAGGGAAGCTGGCCGC 1914  
Db 661 ValLeuArgGlnGlnTyrGlnSerProSerArgGluAlaGlyArg 676

## RESULT 8

US-10-441-949-38  
; Sequence 38, Application US/10441949  
; Publication No. US20040116684A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodrigo, Allen  
; APPLICANT: Ross, Howard A.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES  
; FILE REFERENCE: 08987-012001  
; CURRENT APPLICATION NUMBER: US/10/441,949  
; PRIOR FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: PCT/US01/05288  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/183,659  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 38  
; LENGTH: 678  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially generated peptide  
US-10-441-949-38

Alignment Scores:  
Pred. No.: 8,02e-303 Length: 678  
Score: 3318.00 Matches: 634  
Percent Similarity: 93.9% Conservative: 1  
Best Local Similarity: 93.8% Mismatches: 3  
Query Match: 95.7% Indels: 38  
DB: 4 Gaps: 7

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-10-441-949-38 (1-678)

Qy 1 ATGCATCCCAAGCTTAACCGCGCCACCTCCCGATTCGGGGTGGAAAGCGAAAGACTG 60  
Db 1 MetHisProThrLeuSerArgHisLeuProIleArgGlyLysProLysArgLeu 20  
Qy 61 AANAATCCCTTAGCTTCGCTCCATCGGTGTTCTTACTCTGTCAATAAATCTCTCAG 120  
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrSerGln 40



; LENGTH: 653  
 ; TYPE: PRT  
 ; ORGANISM: Viral  
 US-10-029-656-2

## Alignment Scores:

Pred. No.: 8,83e-287 Length: 653  
 Score: 3147.50 Matches: 593  
 Percent Similarity: 93.6% Conservative: 18  
 Best Local Similarity: 90.8% Mismatches: 27  
 Query Match: 90.8% Indels: 15  
 DB: 4 Gaps: 3

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-10-029-656-2 (1-653)

```

QY 1 ATGCATCCACGTTAAACCGCGCCACCTCCCGATTCCGGGTGGAAGCCGAAAGACTG 60
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 MethHisProThrLeuSerArgArgHisLeuProIleArgGlyLysProIleArgLeu 20
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 61 AAATCCCTTAAGCTTCGCTCCATCGCGTTCCTTACTCTCTCAATAACCTCTCAG 120
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 21 LysIleProLeuSerPheAlaSerIleAlaIlePheLeuThrLeuSerIleThrProGln 40
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 121 ACTAATGCTATCGCATAGGACAGACCTGAACTCCCATAAACCTTATCTCTACCTGG 180
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 41 ValAsnGlyLysArgLeuValAsnSerProAsnSerHisLysProLeuSerLeuThrTrp 60
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 181 TTAATTACTGACTCGGCACAGGTATTAAATCAACAACACTCAAGGGGAGCTCTTTA 240
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 241 GGAACTGGTGGCTGTATATACGTTTTCCTTCAGATCAGATTATTCCTAGTCTG----- 294
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 295 ---ACCTCACCCCGATATCTCCATGCTCAGGATTTTATGTTTGGCCAGGACCA 351
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 352 AATAATGGAACATTGCGGAAATCCAGAGATTCTTTTGTAAACATGGAATGTGTA 411
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 121 AsnAsnGluGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 412 ACCTCTAATGATGATATTGAAATGGCCAACTCTCAGCAGGATAGGTTAAGTTTCT 471
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnAspArgValSerTyrSer 160
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 472 TATGTCACACCTATACCACTCTGGACAATTATATAC-----CTGACC 516
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 517 TGG-----ATTAGAACTGGAAGCCCAAGTCTCTCTCTCAGAC 555
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 556 CTAGATTACCTAAAAATAAGTTTCACTGAGAAGGAAACCAAGAAAATATCTTAAATGG 615
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGlnIleLeuLysTrp 220
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 616 GTAATGTATGCTCTGGGGAATGGTATATTATGAGAGCTCGGTAAACAACACCGCTCC 675
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 221 ValHisGlyMetSerTrpGlyMetValTyrTyrGlyGlySerGlyLysGlnProGlySer 240
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 676 ATTCCTAATATTCCTCAAAATAAACCGCTGAGCTCCCAATGGCTATAGGACCAAT 735
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 241 IleLeuThrIleArgLeuLysIleAsnGlnLeuGluProPheMetAlaIleGlyProAsn 260
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 736 ACGGCTTTGACGGTCAAAGACCCCAACCAAGGACGACCATCTCTTAACATAACT 795
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 261 ThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyThrSerSerAsnIleThr 280
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 796 TCTGGATCAGACCCCACTGAGTCTAACAGCAGCACTAAATGGGGGCAAACTTTTATGC 855
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
  
```

```

Db 281 SerGlySerAspProThrGluSerAsnSerThrThrLysMetGlyAlaLysLeuPheSer 300
QY 856 CTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCAGACTCCAGAGSCTACCTCTTCTGT 915
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 301 LeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThrSerSerCys 320
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 916 TGGCTATGCTTAGCTTCGGGCCACCTTACTATGAAGGAATGGCTAGAGAGGGAATTC 975
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 321 TrpLeuCysLeuAlaLeuGlyProProTyrTyrGluGlyMetAlaArgArgGlyLysPhe 340
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 976 AATGTGACAAAAGAACAATAGACCAATGCATCGGGATCCCAAAAATAAGCTTACCTTT 1035
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 341 AsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGlnAsnLysLeuThrLeu 360
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1036 ACTGAGGTTTCTGGAAGAGGACCTGCATAGAGAAAGTTCCCCATCCCAACACACCTT 1095
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 361 ThrGluValSerGlyLysGlyThrCysIleGlyLysValProProSerHisGlnHisLeu 380
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1096 TGTAAACCACTGAAGCCCTTTAATCAAACTCTGAGAGTCAATATCTGGTACCTGGTTAT 1155
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 381 CysAsnHisThrGluAlaPheAsnGlnThrSerGluSerGlnTyrLeuValProGlyTyr 400
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1156 GACAGTGGTGGCGATGTAATACTGGATAACCCCTTGTGTTCACCTTGGTTTTTAAC 1215
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 401 AspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeuValPheAsn 420
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1216 CAAACTAAAGATTTCGATTTATGTCCTCAAAATGTTCCCGAGTGTATTACTATCCCGAA 1275
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 421 GlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrTyrProGlu 440
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1276 AAAGCAATCTCTGATGTAATATGACTACAGAAATCATCGACAAAAGAGAGAACCCATATCT 1335
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 441 LysAlaIleLeuAspGluTyrAspTyrArgAsnHisArgGlnLysArgGluProIleSer 460
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1336 CTGACACTGCTGTGATGCTCGGACTTGGAGTGGCAGCAGGTAGGAAACAGGAACAGCT 1395
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 461 LeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyValGlyThrAla 480
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1396 GCCCTGTGTCGGGACCAACAGCAGCTAGAAACAGGACTTACTACCTACATCGAATCTTA 1455
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 481 AlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHisArgIleVal 500
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1456 ACAGAAGATCTCCAAGCCCTAGAAAAATCTCTCAGTAACCTGGAGGAATCCCTAACCTCC 1515
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 501 ThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluSerLeuThrSer 520
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1516 TTATCTGAAGTAGTCTCTACAGATAGAGAGGGTTAGATTATTATTCTTAAAGAAGGA 1575
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 521 LeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuLysGluGly 540
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1576 GGATTATGTAGCTTGAAGGAGGAATGCTGTTTTTATGTGGATCATTCAGGGGCCATC 1635
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 541 GlyLeuCysValAlaLeuLysGluGlyCysPheTyrValAspHisSerGlyAlaIle 560
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1636 AGACACTCCATGAACAACTTAGAAGAGGTGGAGAAGCGCTCGAAGGAAAAAGAAACT 1695
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 561 ArgAspSerMetSerLysLeuArgGluArgLeuGluLysArgArgArgGluLysGluThr 580
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1696 ACTCAAGGGTGGTTGAGGATGGTTCAACAGGTCTCTTTGGTTGGCTTACCTACTTCT 1755
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 581 ThrGlnGlyTrpPheGluGlyTyrPheAsnArgSerLeuTrpLeuAlaThrLeuLeuSer 600
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1756 GCTTTACAGGACCTTTAATAGTCTCTCTCTGTACTACAGTTTGGCCATGTATTATT 1815
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 601 AlaLeuThrGlyProLeuIleValLeuLeuLeuLeuLeuThrValGlyProCysIleIle 620
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1816 AACAAAGTTAATGGCTCTTATTAGAGAAGCAATAGTCAGTCCAGATCATCGTACTTAGA 1875
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 621 AsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMetValLeuArg 640
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 1876 CAACAGTACCAAGCCCGTCTAGCAGGGAAGCTGGCCGC 1914
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 641 GlnGlnTyrGlnSerProSerArgSerArgGluAlaGlyArg 653
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
  
```



```
RESULT 10
US-10-029-656-4
; Sequence 4, Application US/10029656
; Publication No. US20030175904A1
; GENERAL INFORMATION:
; APPLICANT: Patience, Clive
; APPLICANT: Oldmixon, Beth
; APPLICANT: Ericsson, Thomas
; TITLE OF INVENTION: Molecular Sequence of Pig Endogenous Retrovirus Receptor and Meth
; FILE REFERENCE: 329579-3
; CURRENT APPLICATION NUMBER: US/10/029,656
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/60/285,103
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Viral
US-10-029-656-4

Alignment Scores:
Pred. No.: 1,196-285 Length: 653
Score: 3135.50 Matches: 591
Percent Similarity: 93.1% Conservative: 17
Best Local Similarity: 90.5% Mismatches: 30
Query Match: 90.4% Indels: 15
DB: 4 Gaps: 3

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-10-029-656-4 (1-653)
QY 1 ATGCATCCCGCTAAACCGCGCCACCTCCGATTCGGGGTGGAAAGCCGAAAGACTG 60
DB 1 MethisProThrLeuSerArgArgHisLeuProIleArgGlyGlyLeuProLysArgLeu 20
QY 61 AAATCCCTTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCAATACCTCTCAG 120
DB 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
QY 121 ACTAATGTATGCGCATAGGACAGACCTGAACTCCCATAAACCTTATCTCTCACCTGG 180
DB 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTAAGTCTCGGCACAGGTATTAAATATCAACACACTCAAGGGAGGCTCCTTTA 240
DB 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
QY 241 GGAACCTGGTGGCTGATCTATACGTTTGGCTCAGATCAGTTATTCCTAGTCTG----- 294
DB 81 GlyThrTrpTrpProGluLeuTrpValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY 295 ---ACCTCACCCAGATATCTCCATCTCAGGATTTTATGTTGTCAGGACCCACCA 351
DB 101 GlnAlaThrProProAspValArgAlaTrpPheTrpValCysProGlyProPro 120
QY 352 AATAATGAAACATTCGCGAAATCCAGAGATTTCTTTTGTAAACAATGGAACCTGTGTA 411
DB 121 AsnAsnGluGluTyrcysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
QY 412 ACCTCTAATGATGGATATTTGGAATGGCCACCTCTCAGCAGGATAGGGTAAGTTTCT 471
DB 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTrpSer 160
QY 472 TATGTCACACCTATACCAGCTCTCGACAATTTTAATTAC-----CTGACC 516
DB 161 PheValAsnAsnProThrSerTrpAsnGlnPheAsnTrpGlyHisGlyArgTrpLysAsp 180
QY 517 TGG-----ATTAGAACTGGAAGCCCCCAAGTGCCTCTCTTCAGAC 555
DB 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
```

```
QY 556 CTAGATTACCTAAAAATAAGTTTCACTGAGAAAGGAAAAACAAGAAAAATATCTTAAATGG 615
DB 201 LeuAspTrpLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleLeuLysTrp 220
QY 616 GTAAATGGTATGTCTTGGGGAATGGTATATTATGGAGGCTCGGTAAACAACACAGGCTCC 675
DB 221 ValAsnGlyMetSerTrpGlyMetValTyrcysGlyGlySerGlyLysGlnProGlySer 240
QY 676 ATTCTAATCTATTCCTCAAAAATAAACACAGCTGAGGCTCAATGGCTATAGGACCAAT 735
DB 241 IleLeuThrIleArgLeuLysIleAsnGlnLeuGluProMetAlaIleGlyProAsn 260
QY 736 ACGGTCTTGGCGGTCAAAGACCCCAACCAAGCAGGACCATCTCTTAAACATAACT 795
DB 261 ThrValLeuThrGlyGlnArgProProThrGlnGlyProGlyProSerSerAsnIleThr 280
QY 796 TCTGGATCAGACCCCACTAGTCTTAACAGCAGCAGCTAAATGGGGGCAAACTTTTACG 855
DB 281 SerGlySerAspProThrGluSerAsnSerThrThrLysMetGlyAlaLysLeuPheSer 300
QY 856 CTATCCAGGAGGCTTTTCAAGCTCTTAACTCCAGCTCCAGAGGCTACCTCTCTTGT 915
DB 301 LeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThrSerSerCys 320
QY 916 TGGCTATCTTCTAGCTTCGGGCCCACTTACTATCAAGGAATGGCTAGAGAGGGAAATTC 975
DB 321 TrpLeuCysLeuAlaLeuGlyProProTyrcysGlyMetAlaArgArgGlyLysPhe 340
QY 976 AATGTGCAAAAGAACATAGAGACCAATGCACATGGGGATCCCAAAATAACCTTACCCCT 1035
DB 341 AsnValThrLysGlyHisArgAspProCysThrTrpGlySerGlnAsnLysLeuThrLeu 360
QY 1036 ACTGAGGTTTCTGAAAGAGGCACCTGCATAGGAAGGTTCCCTCCATCCCAACACACTT 1095
DB 361 ThrGluValPheGlyLysGlyThrCysIleGlyLysValProProSerHisGlnHisLeu 380
QY 1096 TGTAACCACTGAAAGCCTTTAATCAAACTCTCAGAGTCAATATCTGGTACCTGGTTAT 1155
DB 381 CysAsnHisThrGluAlaPheAsnArgThrSerGluSerGlnTyrcysLeuValProGlyTyrc 400
QY 1156 GACAGGTGGTGGGCATGTAACTACCTGATTAACCCCTTGTGTTCACCTCGTGGTTTAA 1215
DB 401 AspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeuValPheAsn 420
QY 1216 CAAACTAAGATTTTGGATTTATGTCCTCAAAATTTGTTCCCGAGTGTATTACTATCCCGAA 1275
DB 421 GlnThrLysAspPheCysIleMetValGlnIleValProArgValTyrcysTrpProGlu 440
QY 1276 AAAGCAATCCTTGATGATATGACTACAGAATCATCGACAAAAGAGAGAACCCATATCT 1335
DB 441 LysAlaIleLeuAspGluTyrcysArgAsnHisArgGlnLysArgGluProLysSer 460
QY 1336 CTGACACTTGTGTGATGCTCGGACTTCGAGTGGCAGCAGGTGTAGGAAACAGGACAGCT 1395
DB 461 LeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGlyValGlyThrGlyThrAla 480
QY 1396 GCCTGTGTCAGGACCCACAGCAGCTAGAAACAGAGACTTAGTAACTCATCTGAATTTGTA 1455
DB 481 AlaLeuValThrGlyProGlnGlnLeuGluThrGlyLeuSerAsnLeuHisArgIleVal 500
QY 1456 ACAGAGATCTCCAGGCCCTAGAAAATCTCTCAGTACCTGGAGGAATCCCTAACCTCC 1515
DB 501 ThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluGluSerLeuThrSer 520
QY 1516 TTATCTCAAGTAGTCTCTACAGAATAGAGGGGTAGATTTTATTATTATTTCTAAAAGAAGA 1575
DB 521 LeuSerGluValValLeuGlnAsnArgGlyLeuAspLeuLeuPheLeuLysGluGly 540
QY 1576 GGATTTATGTGTAGGCTTGAAGGAGGAATGCTGTTTTTATGTGGATCATTTACGGGGCCATC 1635
DB 541 GlyLeuCysValAlaLeuLysGluGluCysCysPheTyrcysPheValAspHisSerGlyAlaIle 560
```



```
QY 1636 AGAGACTCCATGAACAGCTTAGAGAAAGGTTGGAGAGCGTCGAGGGAAGAAAGAACT 1695
Db 561 ArgAspSerMetSerLysLeuArgGluLeuArgGluLysArgGluLysGluThr 580
QY 1696 ACTCAAGGGTGTGGAGGATGTTCAACAGGCTCTTTGGTGGCTACCTCTTCT 1755
Db 581 ThrGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpLeuAlaThrLeuLeuSer 600
QY 1756 GCTTTAAGCAGGACCTTAATAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1815
Db 601 AlaLeuThrGlyProLeuLeuValLeuLeuLeuLeuLeuLeuLeuValGlyProCysIleIle 620
QY 1816 AACAGTAATGCTTCCTCAITAGAGAACGAATAAGTCAGTCCAGATCATGCTTCTAGA 1875
Db 621 AsnLysLeuIleAlaPheIleArgGluArgIleSerAlaValGlnIleMetValLeuArg 640
QY 1876 CAACAGTACCAAGCCCGCTAGCAGGGAAGCTGCCGC 1914
Db 641 GlnGlnTrpGlnSerProSerArgGluAlaGlyArg 653

RESULT 11
US-10-441-949-43
; Sequence 43, Application US/10441949
; Publication No. US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodrigo, Allen
; APPLICANT: Ross, Howard A.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-43

Alignment Scores:
Pred. No.: 1,49e-261 Length: 661
Score: 2879.50 Matches: 541
Percent Similarity: 88.3% Conservative: 40
Best Local Similarity: 82.2% Mismatches: 56
Query Match: 83.0% Indels: 21
DB: 4 Gaps: 6

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-10-441-949-43 (1-661)
QY 1 ATGCATCCCAAGTTAAACCGCGCCACCTCCCGATTCGGGTGGGAAAGCGAAAGACTG 60
Db 1 MetHisProThrLeuSerArgGluLeuProIleArgGlyLysProLysArgLeu 20
QY 61 AAAATCCCTTAGCTTCGCTTCATCGGTGGTCTTACTCTGTCAATAACTCTCTCAG 120
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
QY 121 ACTAATGGTATGCGCATAGGAGCAGCTGAACTCCCATAAACCTTATCTCTCAGCTGG 180
Db 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTACTGACTCCCGCAGCAGGATTATTAATCAACAACACTCAAGGGAGGCTCCCTTA 240
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
QY 241 GSAACCTGGTGGCTGTATCTATACGTTTGGCTCAGATCAGTTATTCCTAGTCTG----- 294
```

```
Db 81 GlyThrTrpTrpProGluLeuLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY 295 ---ACCTCACCCAGATATCTCCATGCTCAGGATTTTATGTTGTCGCCAGGACCACA 351
Db 101 GlnAlaThrProProAspValLeuAlaThrGlyPheTyrValCysProGlyProPro 120
QY 352 AATAATGGAAACATTCGGAAATCCCGAGATTTCTTTTGTAAACAATGAACATGCTGA 411
Db 121 AsnAsnGluGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
QY 412 ACCTCTAATGATGATATTTGGAATGCGCAACCTCTCAGCAGGATAGGTTAAGTTTCT 471
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160
QY 472 TATGCTCAACACCTATACAGCTCTGGACAATTTAATTAC-----CTGACC 516
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
QY 517 TGG-----ATTAGAACTGGAAGCCCAAGTCTCTCTTCACGAC 555
Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysAsnSerLeuAsp 200
QY 556 CTGATTACCTAAATAAGTTTCTCAGTGAAGAAAGAAACAAGAAAATATATCTTAAATGG 615
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrp 220
QY 616 GTAATGATGATGCTTCGGGAATGGTATATATGAGGCTCGGCTGAACAACACCGCTCC 675
Db 221 ValAsnGlyMetSerTrpGlyIleValTyrTrpGlySerGlyArgLysGlyLysTrp 240
QY 676 ATTCTAACTATTCGCTCAAAATA---AACCAGCTGAGGCTCCCAATGGCTATAGGACCA 732
Db 241 ValLeuThrIleArgLeuArgGlyGluThrGlnMetGluProProValAlaIleGlyPro 260
QY 733 AATACGGTCTTGACGGGTCAAGACCCCAACCCAA-----GGACAGGACCA 780
Db 261 AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGluArgProSerProAsnPro 280
QY 781 TCCTCT---AACATACTTCTGGATCAGACCCCACTAGTCTACACGACGACTAAAATG 837
Db 281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300
QY 838 GGGGCAAACTTTTAGCTCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCACGACTCCA 897
Db 301 GlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320
QY 898 GAGGCTACCTCTTCTTGTGGCTATGCTTAGCTTCGGGCCCCACCTTACTATGAAGGAATG 957
Db 321 GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMet 340
QY 958 GCTAGAAGAGGGAATTCATGTGACAAAAGAACATAGACACCAATGCATGGGGATCC 1017
Db 341 AlaArgGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySer 360
QY 1018 CAAAATAAGCTTACCTTACTGAGGTTTCTGGAAGGCACTGCATAGAAAGGTTCCC 1077
Db 361 GlnAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyArgValPro 380
QY 1078 CCATCCCAACCAACCTTTGTAAACACACTGAAGCCCTTAAATCAAACTCTGAGAGTCAA 1137
Db 381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400
QY 1138 TATCTGCTACCTGCTTATGACAGGTGGTGGCATGTAATACTGGATTAACCCCTTGTT 1197
Db 401 TyrLeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVal 420
QY 1198 TCCACCTTGTTTAAACCAACTAAAGATTTTGCATTATGTCCTAAATTTGTCCTCCCA 1257
Db 421 SerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArg 440
QY 1258 GTGTATTACTATCCCGAAAGCAATCCTTGTGATGAATATGACTACAGAAATCATCGCAA 1317
```

```
Db 441 ValTyrTyrProGluLysAlaValLeuAspGluTyrAspTyrArgTyrAsnArgPro 460
Qy 1318 AAGAGAACCCATATCTCTGACACTTCTGTGATGCTCGCACTTGGAGTGGCAGCAGGT 1377
Db 461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGly 480
Qy 1378 GTAGAAACAGAACAGCTGCTGTCTACGGGACCACAGCAGCTAGAAACAGGACTTAGT 1437
Db 481 ValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSer 500
Qy 1438 AACCTACATCGAATTGTAAACAAGAGACTCTCAAGCCCTAGAAAAATCTGTCTAGTAACCTG 1497
Db 501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520
Qy 1498 GAGGAATCCCTAACTCTTATCTGAAGTAGTCTCTACAGATAGAAAGGTTAGATTTA 1557
Db 521 GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 540
Qy 1558 TTATTCTTAAAGAGGAGGATTATGTGTAGCTTGAAGGAGGAATGCTGTTTTTATGTG 1617
Db 541 LeuPheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysPheTyrVal 560
Qy 1618 GATCATTCAGGGCCCATCAGAGACTCCATGAACAGCTTAGAGAAAGCTTGGAAAGCGT 1677
Db 561 AspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArg 580
Qy 1678 CGAAGGAAAGGAAACTACTCAAGGTGGTTTGAAGGATGGTTCAACAGTCTCTTTGG 1737
Db 581 ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrp 600
Qy 1738 TTGGTCACTTCTTCTGCTTAAACAGCCCTTAATAGTCCCTCTCTGTACTCTACA 1797
Db 601 MetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThr 620
Qy 1798 GTTGGCCCATGTATTATTAACAAGTTAATGTCTTTCATTAGAGAACGAATAAGTGCAGTC 1857
Db 621 ValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaVal 640
Qy 1858 CAGATCATGGTACTTAGACACAGTACCAAGCCCGTCTAGCAGGAAGCTGGC 1911
Db 641 GlnIleMetValLeuArgGlnGlnTyrGlnGlyLeuLeuSerGlnGlyGluGly 658
```

## RESULT 12

```
US-10-441-949-16
; Sequence 16, Application US/10441949
; Publication No. US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodrigo, Allen
; APPLICANT: Rose, Howard A.
; APPLICANT: Mullins, James I.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; PRIOR FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-16
```

```
Alignment Scores:
Pred. No.: 4.4e-261 Length: 660
Score: 2874.50 Matches: 539
Percent Similarity: 88.8% Conservative: 38
Best Local Similarity: 82.9% Mismatches: 52
```

```
Query Match: 82.9% Indels: 21
DB: 4 Gaps: 6
US-10-723-552-3_copy_5620_7533 (1-1914) x US-10-441-949-16 (1-660)
Qy 1 ATGCATCCCAAGTTAAACCGCGCCACCTCCGATTCGGGTGGAAGCCGAAAGACTG 60
Db 1 MethiSprothrLeuSerArgArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20
Qy 61 AAAATCCCTTAAGCTTCGCTCCATCGCGTGGTTCCTTACTCTGTCAATTAACCTCTCAG 120
Db 21 LysileProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
Qy 121 ACTAATCGTATGCCATAGGACAGCCTGAACCTCCATAAACCTTATCTCTCACCTGG 180
Db 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
Qy 181 TTAATTACTGACTCCGCGCACAGGTATTAAATATCAACAACTCAAGGGAGGCTCTTTTA 240
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
Qy 241 GGAACCTGGTGGCTGTATACGTTCCTCAGATCAGTATTCTTAGTCTG----- 294
Db 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
Qy 295 ---ACCTCACCCAGATATCTCCTCATGCTCACGATTTTATGTTTGGCCAGGACCACA 351
Db 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
Qy 352 AATAATGAAACCAATTGCGGAAATCCCAGAGATTCTTTTGTAAACAATGGAACGTGTA 411
Db 121 AsnAsnGluGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
Qy 412 ACCTCTAATGATGATNTGGAAATGCGCAACCTCTCAGCAGGATAGGTTAAGTTTCT 471
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpValSerGlnGlnAspArgValSerTyrSer 160
Qy 472 TATGTCACACCTATACCAGCTCTGGACAAATTTAATTAC-----CTGACC 516
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
Qy 517 TGG-----ATTAGAACTCGAAGCCCAAGTGTCTCTCTTCAGAC 555
Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
Qy 556 CTAGATTACCTAAATAAGTTTCACTGAGAAAGGAAACAGAAATAATCTCTAAATGG 615
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrp 220
Qy 616 GTAAATGGTATGCTTGGGGAATGTTATATTATGGAGCTCGGTAAACAAACACGAGCTCC 675
Db 221 ValAsnGlyMetSerTrpGlyIleValTyrTyrGlyGlySerGlyArgLysLysGlySer 240
Qy 676 ATTCTAATCTATTCGCTCAAAATA---AACAGCTGGAGCCTCCCAATGGCTATAGGACCA 732
Db 241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProProValAlaIleGlyPro 260
Qy 733 AATACGGTCTTGACGGGTCAAGAGCCCAACCCNA-----GGACCAGGACCA 780
Db 261 AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGluGlnArgProSerProAsnPro 280
Qy 781 TCCTCT---AACATAACTTCTGGATCAGACCCCACTGAGTCTAACAGCAGCACTAAATG 837
Db 281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300
Qy 838 GGGGCAAACTTTTATAGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCTCAACGACTCCA 897
Db 301 GlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320
Qy 898 GAGGTACTCTTCTTGTGGCTATGCTTACGTTCGGGCCCACTTACTATATGAAGGAATG 957
Db 321 GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMet 340
```

```
QY 958 GCTAGAGAGCGAAATTCATGTGACAAAAGACATAGACACCAATGCACATGGGATCC 1017
Db 341 AlaArgGlyGlySerPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySer 360
QY 1018 CAAATAAGCTTACCTTACTGAGGTTCTCGAAAGGCACCTGCATAGAAAGGTTCCC 1077
Db 361 GlnAsnLysLeuThrLeuGluValSerGlyLysGlyThrCysIleGlyArgValPro 380
QY 1078 CCATCCCAACCAACCTTTGTAACCAACACTGAAGCCTTTAATCAAAACCTCTGAGAGTCAA 1137
Db 381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400
QY 1138 TATCTGGTACCTGTTATGACAGGTGGTGGCATGTATATCTGAGTAACCCCTTGTTGT 1197
Db 401 TyrLeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVal 420
QY 1198 TCCACCTTGGTTTAAACCAACTAAAGATTTTGCATTATGTCCTCAAAATGTTCCCGA 1257
Db 421 SerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArg 440
QY 1258 GTGTATTACTATCCGAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAA 1317
Db 441 ValTyrTyrTrpProGluLysAlaValLeuAspGluTyrAspTyrArgTyrAsnArgPro 460
QY 1318 AAGAGAGAACCATATCTCTGACACTTCTGTGATGCTCGGACTTGGAGTGGCAGCAGGT 1377
Db 461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGly 480
QY 1378 GTAGGACAGGAACAGCTGCCCTCGTCACGGGACACAGCAGCTAGAAACAGGACTTAGT 1437
Db 481 ValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSer 500
QY 1438 AACCTACATCGAATTTGTAACAGAAAGATCTCCAAGCCCTAGAAAAATCTGTCAAGTAACTG 1497
Db 501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520
QY 1498 GAGGAATCCCTAACCTCTTATCTGAAGTAGTCTCTACAGAAAGAGGTTAGATTTA 1557
Db 521 GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 540
QY 1558 TTATTTCTAAAGAGAGGAGGATTATGTAGCCTTGAAGGAGGAGTCTGTTTTTATGTG 1617
Db 541 LeuPheLeuLysGluGlyGlyLeuCysValAlaLeuLysGluGluCysPheTyrVal 560
QY 1618 GATCATTCAGGGGCCATCAGAGACTCCATGAACAAAGCTTAGAGAAAGGTTGGAGAAGCGT 1677
Db 561 AspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArgArg 580
QY 1678 CGAAGGAAAGGAAACTACTCAAGGTGGTTTGAGGGATGGTTCAACAGGTCTCTTTGG 1737
Db 581 ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrp 600
QY 1738 TTGGCTACCTACTTCTGCTTTAACAGGACCTTAATAGTCTCTCTCTCTCTCTCTCTCT 1797
Db 601 MetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThr 620
QY 1798 GTTGGGCCATGTATTATTAAACAGTAAATGTCCTTCAATTAGAGAACCAATAAGTGCAGTC 1857
Db 621 ValGlyProCysLeuLeuLeuAsnArgPheValAlaPheValArgGluArgValSerAlaVal 640
QY 1858 CAGATCATGTTACTTAGACAACTACCA 1887
Db 641 GlnIleMetValLeuArgGlnGlnTyrGln 650
```

RESULT 13

```
US-10-441-949-8
; Sequence 8, Application US/10441949
; Publication No. US20040116684A1
; GENERAL INFORMATION:
; APPLICANT: Rodrigo, Allen
; APPLICANT: Ross, Howard A.
; APPLICANT: Mullins, James I.
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES
```

```
; FILE REFERENCE: 08987-012001
; CURRENT APPLICATION NUMBER: US/10/441,949
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: PCT/US01/05288
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/183,659
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Artificially generated peptide
US-10-441-949-8
```

```
Alignment Scores:
Pred. No.: 8,43e-261 Length: 660
Score: 2871.50 Matches: 539
Percent Similarity: 88.6% Conservative: 37
Best Local Similarity: 82.9% Mismatches: 53
Query Match: 82.8% Indels: 21
DB: 4 Gaps: 6
```

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-10-441-949-8 (1-660)

```
QY 1 ATGCATCCCAAGTTAAACCGCGCCACTCCCGATTCGGGGTGGAAAGCCGAAAGACTG 60
Db 1 MethisProthrLeuSerArgArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20
QY 61 AAAATCCCTTAAAGCTTCGCTCCATCGCTGGTTCCTTACTCTCTCAATAACCTCTCAG 120
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
QY 121 ACTAATGGTATGCGCATAGGAGCAGCTGAACTCCCATAAACCTTATCTCTCACCTGG 180
Db 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTAATCTACCTCCGSCACAGGTATTATATCAACAACACTCAAGGGAGGCTCTTTA 240
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80
QY 241 GGAACCTGTGTGGCTGTATCTATATACGTTTGCCTCAGATCAGTATTTCCTAGTCTG 294
Db 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY 295 ---ACCTCACCCAGATATCTCCATGCTCAGGATTTTATGTTGCCAGGACCA 351
Db 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
QY 352 AATAATGGAACCAATTCGGAATCCAGAGATTTCTTTTAAACAATCGAACTGTGTA 411
Db 121 AsnAsnGluGlyTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
QY 412 ACCTCTAATCATGGATATTGGAATGCGCAACTCTCAGCAGGATAGGGTAAGTTTCT 471
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160
QY 472 TATGTCAACACCTTATACAGCTCTGGACAATTTAATTAC-----CTGACC 516
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
QY 517 TGG-----ATTAGACTGGAAGCCCAAGTCTCTCTCTCTCTCTCTCTCT 555
Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
QY 556 CTAGATTACCTAAATAAGTTTCACTGAGAAAGGAAACACAGAAATATATCTTAAATGG 615
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyGlnGlnAsnIleGlnLysTrp 220
QY 616 GTAATATGGTATGCTTTGGGAATGGTATATTATGGAGGCTCGGTAACCAACCGGCTCC 675
```

Db 221 ValAsnGlyMetSerTrpGlyLeuValTyrTyrGlyGlySerGlyArgLysLysGlySer 240  
Qy 676 ATTCTAACTATTCGCTCAAAATA---AACAGCTGGAGCCTCCAAATGGCTATAGGACCA 732  
Db 241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProValAlaIleGlyPro 260  
Qy 733 AATACGGTCTTGACGGGTCAAGACCCCAACCCAA-----GGACACAGGACCA 780  
Db 261 AsnLysGlyLeuAlaGluGlnGlyProIleGlnGluArgProSerProAsnPro 280  
Qy 781 TCCTCT---AACATAACTTCTGGATCAGACCCCACTGAGTCTAAACAGCAGCACTAAATG 837  
Db 281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300  
Qy 838 GGGCAAAACTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAAGTCTTAAGTCTCAACACTCCA 897  
Db 301 GlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320  
Qy 898 GAGGTACTCTCTTGTGTGGCTATGCTTACGTTTCGGGCCCACTTACTATGAAGGAATG 957  
Db 321 GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMet 340  
Qy 958 GCTAGAAGAGGAATAATGTGACAAAAGAACATAGACCAATGACATGGGATGCC 1017  
Db 341 AlaArgGlyGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySer 360  
Qy 1018 CAAATAAGCTTACCTTACTGAGTTTCTCGAAAGGCACCTGCATAGGAAGGTTCCTCC 1077  
Db 361 GlnAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyMetValPro 380  
Qy 1078 CCATCCCAACCAACCTTTGTAACACCACTGAAGCTTTAATCAACCTCTGAGAGTCAA 1137  
Db 381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400  
Qy 1138 TATCTGTGTACCTGGTTATGACAGGTGGTGGGCATGTATATCTGGATTAACCCCTGTGTT 1197  
Db 401 TyrLeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVal 420  
Qy 1198 TCGACCTTGGTTTAAACCAACTTAAGATTTTGCATTATGGTCCAAATTTGTTCCCGGA 1257  
Db 421 SerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArg 440  
Qy 1258 GTGTATTACTATCCGAAAAAGCACTCTCATGAATATGACTACAGAAATCATCGCAA 1317  
Db 441 ValTyrTyrTyrProGluLysAlaValLeuAspGluTyrAspTyrArgTyrAsnArgPro 460  
Qy 1318 AAGAGAGAACCATATCTCTGACACTGCTGTGATGCTCGGACTTGGAGTGGCAGCAGGT 1377  
Db 461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGly 480  
Qy 1378 GTAGAACAGGAACAGCTGCCTGTACGGGACCAAGCAGCTAGAACAGGACTTAGT 1437  
Db 481 ValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSer 500  
Qy 1438 AACCTACATCGAATTTGTAACAGACATCTCAAGCCCTAGAAAAATCTGTCAGTAAACCTG 1497  
Db 501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuLysSerValSerAsnLeu 520  
Qy 1498 GAGGAATCCCTAACCTCTTATCTGAAGTAGTCTTACAGATAGAGAAGGTTTATGATTTA 1557  
Db 521 GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 540  
Qy 1558 TTATTTCTAAAGAGGAGGATTATGTACCTTGAAGGAGGATGCTGTTTTTATGTG 1617  
Db 541 LeuPheLeuLysGluGlyLeuCysValAlaLeuLysGluGluCysPheTyrVal 560  
Qy 1618 GATCATTACAGGGCCATCAGACATCCATGACAAAGCTTTAGAGAAAGGTTTCGAGAAGCGT 1677  
Db 561 AspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArgArg 580  
Qy 1678 CGAAGGGAAGGAACCTACTCAAGGGTGGTTTGAAGGATGTTCAACAGGTCTCTTTGG 1737  
Db 581 ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrp 600

Qy 1738 TTGGCTACCTACTTTCTGCTTTAAACAGGACCCCTTAATAGTCTCTCTCTGTTACTCACA 1797  
Db 601 MetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThr 620  
Qy 1798 GTTGGGCCATGTATTATTAAACAAGTTAATTGCCTTCAATTAGAGAACGAATAAGTGCAGTC 1857  
Db 621 ValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaVal 640  
Qy 1858 CAGATCATGTGTACTTACAGAACAGTACCAA 1887  
Db 641 GlnIleMetValLeuArgGlnGlnTyrGln 650  
RESULT 14  
US-09-851-859A-5  
; Sequence 5, Application US/09851859A  
; Patent No. US20020065407A1  
; GENERAL INFORMATION:  
; APPLICANT: Banerjee, Papia T.  
; APPLICANT: Patience, Clive  
; APPLICANT: Andersson, Goran K.  
; TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of  
; FILE REFERENCE: 61750-321  
; CURRENT APPLICATION NUMBER: US/09/851.859A  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR FILING DATE: 1998-08-18  
; PRIOR APPLICATION NUMBER: US 09/376781  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 660  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PERV-A  
; OTHER INFORMATION: polypeptide sequence taken from GenBank Accession  
; OTHER INFORMATION: No. US20020065407A1 Y12238 for comparison.  
US-09-851-859A-5

## Alignment Scores:

Pred. No.:	2,49e-260	Length:	660
Score:	2866.50	Matches:	537
Percent Similarity:	88.6%	Conservative:	39
Best Local Similarity:	82.6%	Mismatches:	53
Query Match:	82.7%	Indels:	21
DB:	3	Gaps:	6

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-09-851-859A-5 (1-660)

Qy 1 ATGCATCCAGCTTTAAACCGGCGCCACTCCCGATTCCGGGTGGAAAGCCGAAAGACTG 60  
Db 1 MethisProThrLeuSerArgArgHisLeuProIleArgGlyGlyLysProLysArgLeu 20  
Qy 61 AAAATCCCTTAAGCTTCGCTCCATCGCGGTTCCTTACTCTGTGTAATACCTCTCAG 120  
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40  
Qy 121 ACTAATGTGTATGCGCATAGGAGACAGCTGCAACTCCCAATAACCTTTATCTCTCACCTGG 180  
Db 41 ValAsnGlyLysArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60  
Qy 181 TTAATTAAGTCCGCGCACAGGTATTATATCAACACTCAAGGGAGGCTCTTTTA 240  
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyGluAlaProLeu 80  
Qy 241 GGAACCTGGTGGCTGTATATAGCTTTGCTTCAGATCAGTTATTCTTAGTCTG----- 294  
Db 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100  
Qy 295 ---ACCTCACCCCCAGATATCTCCATCGTCTCAGCGATTTTATGTGTTGCCAGGACCA 351

Db 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120  
QY 352 AATAATGGAACAACTTGGGAAATCCAGAGATTCTTTTCTAACAATGGAATGCTGA 411  
Db 121 AsnAsnGluTyrCysGlyAsnProGlnAspPheCysLysGlnTrpSerCysIle 140  
QY 412 ACCTCTAATGATGATATGGAAATGGCCAACTCTCAGCAGGATAGGGTAAGTTTCT 471  
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160  
QY 472 TATGTCAACACTATACAGCTCTGGACAATTTAATTAC-----CTGACC 516  
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180  
QY 517 TGG-----ATTAGAACTGGAAGCCCAAGTGTCTCTCTCCCTCAGAC 555  
Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200  
QY 556 CTAGATTACTTAAATAAAGTTTCACTGAGAAAGGAAACAAAGAAATATATCTAAATGG 615  
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrp 220  
QY 616 GTAAATGGTATGCTTGGGGAATGGTATATTTATGAGGCTCGGTAAACAACAGGCTCC 675  
Db 221 ValAsnGlyIleSerTyrGlyValTyrTyrGlyGlySerGlyArgLysLysGlySer 240  
QY 676 ATTCTAACTATTCCTCCCAATAA---AACCAGCTGGAGCTCCCAATGGCTATAGGACCA 732  
Db 241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProProValAlaIleGlyPro 260  
QY 733 AATACGGCTTTGACGGGTCAAGAGACCCCAACCCAA-----GGACCAAGGACCA 780  
Db 261 AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGlnArgProSerProAsnPro 280  
QY 781 TCCTCT---AACATACTTCTGGATCAGACCCCACTGAGTCTAAGCAGCAGCACTAAATG 837  
Db 281 SerAspTyrAsnThrThrSerGlySerValProThrGluProAsnIleThrIleLysThr 300  
QY 838 GGGGCAAACTTTTAGCCTCATCCAGGAGCTTTCAAGCTCTTAAGCTCCACAGCACTCCA 897  
Db 301 GlyAlaLysLeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrPro 320  
QY 898 GAGCTACCTCTCTCTGCTATGCTTAGCTTCGGGGCCACCTTACTATGAGGAATG 957  
Db 321 GluAlaThrSerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMet 340  
QY 958 GCTAGAAGAGGAAATTCATGTGACAAAAGAACATAGACCAATGCATGGGATCC 1017  
Db 341 AlaArgGlyGlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySer 360  
QY 1018 CAAAATAAGCTTACCTTACTGAGGTTCCTGGAAGGCCACCTGCATAGGAAGTTCCC 1077  
Db 361 GlnAsnLysLeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyMetValPro 380  
QY 1078 CCATCCCAACCAACCTTTGTAACCACTGAAGCTTTAATCAAACTCTGAGAGTCAA 1137  
Db 381 ProSerHisGlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGln 400  
QY 1138 TATCTGTACTGTTATGACAGTGTGGCATGTAAATCTGGAATTAACCCCTTGTT 1197  
Db 401 TyrLeuValProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysVal 420  
QY 1198 TCACCTTGTTTAAACCAACTTAAAGATTTTGCATTATGCTCCAAATTTGTTCCCGA 1257  
Db 421 SerThrLeuValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArg 440  
QY 1258 GTGTATTACTATCCCGAAAAGCAATCTTGTATGAATATGACTACAGAAATCATCGCAA 1317  
Db 441 ValTyrTyrProGluLysAlaValLeuAspGluTyrAspTyrArgTyrAsnArgPro 460  
QY 1318 AAGAGAAACCAATATCTGACACTTGTGTGATGCTCGGACTTGGAGTGGCAGCAGGT 1377

Db 461 LysArgGluProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaAlaGly 480  
QY 1378 GTAGAAACAGGAACAGCTGCCCTGTCACGGGACACAGCAGCTAGAAACAGGACTTAGT 1437  
Db 481 ValGlyThrGlyThrAlaAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSer 500  
QY 1438 AACCTACATCGAATTTGTAACAGAGATCTCCAAGCCCTAGAAAAATCTGTCAAGTAACTG 1497  
Db 501 AsnLeuHisArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeu 520  
QY 1498 GAGAAATCCCTAACCTCTTATCTGAAGTAGTCTCAGATAAGAGGGTTAGATTTA 1557  
Db 521 GluGluSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeu 540  
QY 1558 TTATTTCTAAAAGAGGAGGATTATGTAGCTTGAAGGAGGAAATCTGTTTTTATCTG 1617  
Db 541 LeuPheLeuLysGluGlyGlyCysValAlaLeuLysGluGlyCysPheTyrVal 560  
QY 1618 GATCATTTCAGGGGCCATCAGAGACTCCATGAACAAGCTTGAAGAAAGGTTGGAGAAGCGT 1677  
Db 561 AspHisSerGlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArgArg 580  
QY 1678 CGAAGGAAAGGAAACTACTCAAGGCTGTTGAGGAGTGGTTCACAGGTCTCTTTGG 1737  
Db 581 ArgArgGluArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrp 600  
QY 1738 TTGGCTACCTTCTCTCTGCTTTAAAGAGGACCTTAATAGTCTCTCTCTGTTACTCACA 1797  
Db 601 MetThrThrLeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuLeuThr 620  
QY 1798 GTTGGGCCATGTATTATTAACTTAACTTGCCTTCAATTAGAGAACGAATAAGTGCAGTC 1857  
Db 621 ValGlyProCysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaVal 640  
QY 1858 CAGATCATGTACTTAGACAACAGTACCAA 1887  
Db 641 GlnIleMetValLeuArgGlnGlnTyrGln 650

RESULT 15  
US-10-441-949-12  
; Sequence 12, Application US/10441949  
; Publication No. US20040116684A1  
; GENERAL INFORMATION:  
; APPLICANT: Rodrigo, Allen  
; APPLICANT: Ross, Howard A.  
; APPLICANT: Mullins, James I.  
; TITLE OF INVENTION: ANCESTRAL VIRUSES AND VACCINES  
; FILE REFERENCE: 08987-012001  
; CURRENT APPLICATION NUMBER: US/10/441,949  
; PRIOR FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: PCT/US01/05288  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/183,659  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 678  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially generated peptide  
US-10-441-949-12

Alignment Scores:  
Pred. No.: 7,24e-259 Length: 678  
Score: 2851.00 Matches: 539  
Percent Similarity: 86.5% Conservative: 38  
Best Local Similarity: 80.8% Mismatches: 52  
Query Match: 82.2% Indels: 38  
DB: 4 Gaps: 7

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-10-441-949-12 (1-678)

```
QY 1 ATGCATCCCGCTAAACCGCGCCACCTCCCGATTCTCGGGTGGAAACCGCAAAAGACTG 60
Db 1 MethisProThrLeuSerArgHisIleuProIleArgGlyGlyProIleArgLeu 20
QY 61 AAAATCCCTTAAAGCTTCGCTCCATCCGCTCGGTTCCTTACTCTGTCAATAACCTCTCAG 120
Db 21 LysIleProLeuSerPheAlaSerIleAlaTrpPheLeuThrLeuSerIleThrProGln 40
QY 121 ACTAATGATATGCGCATAGGAGACGCTGAACCTCCCATAAACCTTATCTCTCACCTGG 180
Db 41 ValAsnGlyIleArgLeuValAspSerProAsnSerHisLysProLeuSerLeuThrTrp 60
QY 181 TTAATTAAGTCTCGGACACAGGATTAATATCAACACACTCAAGGGAGGCTCCCTTTA 240
Db 61 LeuLeuThrAspSerGlyThrGlyIleAsnIleAsnSerThrGlnGlyIleValAlaProLeu 80
QY 241 GGAACCTCGTGGCTGATCTATACGTTTGGCTTCAGATCAGTATTTCCTAGTCTG----- 294
Db 81 GlyThrTrpTrpProGluLeuTyrValCysLeuArgSerValIleProGlyLeuAsnAsp 100
QY 295 ---ACCTCACCCCGATATCTCATGCTCAGCGATTTTATGTTGGCCAGGACCA 351
Db 101 GlnAlaThrProProAspValLeuArgAlaTyrGlyPheTyrValCysProGlyProPro 120
QY 352 AATAATGGAACATTCGGAATCCAGAGATTCTTTCTTAACATGGAATGTA 411
Db 121 AsnAsnGluGluTyrCysGlyAsnProGlnAspPhePheCysLysGlnTrpSerCysVal 140
QY 412 ACCTCTAATGATGATATTCGGAATGGCCAACTCTCAGCAGGATAGGTTAAGTTTCT 471
Db 141 ThrSerAsnAspGlyAsnTrpLysTrpProValSerGlnGlnAspArgValSerTyrSer 160
QY 472 TATGTCAACACTATACAGCTCTGGACAATTTAATTAC-----CTGACC 516
Db 161 PheValAsnAsnProThrSerTyrAsnGlnPheAsnTyrGlyHisGlyArgTrpLysAsp 180
QY 517 TGG-----ATTAGACTGGAAGCCCCCAGTCTCTCTTCAGAC 555
Db 181 TrpGlnGlnArgValGlnLysAspValArgAsnLysGlnIleSerCysHisSerLeuAsp 200
QY 556 CTAGATTACCTAAATAAAGTTTCACTGAGAAAGGAAACAAAGAAATATCTTAAATGG 615
Db 201 LeuAspTyrLeuLysIleSerPheThrGluLysGlyLysGlnGluAsnIleGlnLysTrp 220
QY 616 GTAATGTATGCTCTGGGGAATGTATATATGAGGCTCGGTGAACAACACAGGCTCC 675
Db 221 ValAsnGlyMetSerTrpGlyLeValTyrTyrGlyGlySerGlyArgLysLysGlySer 240
QY 676 ATTCTAAGTATTCGCTCAAAATA---AACCAGCTGGAGCTCCCAATGGCTATAGGACCA 732
Db 241 ValLeuThrIleArgLeuArgIleGluThrGlnMetGluProProValAlaIleGlyPro 260
QY 733 AATACGGTCTTGACGGGTCAAGAGACCCCAACCCCAAGGACCA----- 774
Db 261 AsnLysGlyLeuAlaGluGlnGlyProProIleGlnGluProProHisAsnLeuProVal 280
QY 775 -----GGACCACTCTCT-----AACATAACTTCTCGA 801
Db 281 ProGlnArgProSerProAsnProAspIleThrGlnSerAspTyrAsnThrThrSerGly 300
QY 802 TCAGACCCCACT-----GAGTCTAACACACGACTAAATAATGGGGCAAAA 846
Db 301 SerValProThrAsnThrProArgAsnGluProAsnIleThrIleLysThrGlyAlaLys 320
QY 847 CTTTTAGCTCATCCAGGAGCTTTTCAAGCTCTTAACTCTCAACGACTCCAGAGGCTACC 906
Db 321 LeuPheSerLeuIleGlnGlyAlaPheGlnAlaLeuAsnSerThrThrProGluAlaThr 340
QY 907 TCTTCTGTGTGCTATGCTTACGTTTCGGGCCACCTTACTATGAAGGAATGGCTAGAGA 966
Db 341 SerSerCysTrpLeuCysLeuAlaSerGlyProProTyrTyrGluGlyMetAlaArgGly 360
```

```
QY 967 GGGAAATTCATGTGACAAAGAAACATAGAGACCAATGCATCGGGATCCAAAAATAAG 1026
Db 361 GlyLysPheAsnValThrLysGluHisArgAspGlnCysThrTrpGlySerGlnAsnLys 380
QY 1027 CTTAACCTTACTAGGTTTCTGGAAAAGGACCTGCGATAGGAAAAGTTTCCCCCATCCAC 1086
Db 381 LeuThrLeuThrGluValSerGlyLysGlyThrCysIleGlyArgValProProSerHis 400
QY 1087 CAACACCTTTGTACACCACTGAAGCCTTTAATCAAACTCTGAGAGTCAATATCTGTGA 1146
Db 401 GlnHisLeuCysAsnHisThrGluAlaPheAsnArgThrSerGluSerGlnTyrLeuVal 420
QY 1147 CCTGGTTATGACAGCTGGTGGGCATGTAATACTAGATTAAACCCCTGTGTTCACCTTG 1206
Db 421 ProGlyTyrAspArgTrpTrpAlaCysAsnThrGlyLeuThrProCysValSerThrLeu 440
QY 1207 GTTTTAAACCAAACTAAAGATTTTGTGATTATGTCCAAATTTGTTCCTCCGAGTGTATTAC 1266
Db 441 ValPheAsnGlnThrLysAspPheCysValMetValGlnIleValProArgValTyrTyr 460
QY 1267 TATCCCGAAAAGCAATCCTTGATGAATATGACTACAGAAATCATCGACAAAAGAGAGAA 1326
Db 461 TyrProGluLysAlaValLeuAspGluTyrAspTyrArgTyrAsnArgProLysArgGlu 480
QY 1327 CCCATATCTCTGACACTTGTCTGATGCTCGACTTGGAGTGGCAGCAGGTGTAGGAAACA 1386
Db 481 ProIleSerLeuThrLeuAlaValMetLeuGlyLeuGlyValAlaIaGlyValGlyThr 500
QY 1387 GGAACAGCTGCCCTGGTCCGCGGACCAACAGCAGCTAGAAAACAGCACTTAGTAACTACAT 1446
Db 501 GlyThrAlaLeuIleThrGlyProGlnGlnLeuGluLysGlyLeuSerAsnLeuHis 520
QY 1447 CGAATTGTACAGAAAGATCTCCAAGCCCTAGAAAATCTGTCACTAAGCTGGAGGAATCC 1506
Db 521 ArgIleValThrGluAspLeuGlnAlaLeuGluLysSerValSerAsnLeuGluGluSer 540
QY 1507 CTAACCTCTCTATCTGAAGTAGTCTCTACAGAAATAGAGAGGTGTAGATTTATTATTCTA 1566
Db 541 LeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuPheLeu 560
QY 1567 AAAGAAGAGGATTATGTGTAGCCTTGAAGAGGAATGCTGTTTTTATGTGGATCATTTCA 1626
Db 561 LysGluGlyGlyLeuCysValAlaLeuLysGluCysCysPheTyrValAspHisSer 580
QY 1627 GGGGCCATCAGAGACTCCATGAACAAGCTTAGAAAAGTTGGAGAACGCGTCGAGGGAA 1686
Db 581 GlyAlaIleArgAspSerMetSerLysLeuArgGluArgLeuGluArgArgArgGlu 600
QY 1687 AAGGAAACTACTCAAGGTGTTTGGAGGATGGTTTCAACAGGTCTCTTTGTTGGCTACC 1746
Db 601 ArgGluAlaAspGlnGlyTrpPheGluGlyTrpPheAsnArgSerProTrpMetThrThr 620
QY 1747 CTACTTTCTGCTTTAACAGGACCTTAAATAGTCTCTCTGTTACTCACAAGTTGGGCCA 1806
Db 621 LeuLeuSerAlaLeuThrGlyProLeuValValLeuLeuLeuLeuThrValGlyPro 640
QY 1807 TGTATTATTAAACAGTTAAATTCCTTCAATTAGAGACGAATAAGTGCAGTCCAGATCATG 1866
Db 641 CysLeuIleAsnArgPheValAlaPheValArgGluArgValSerAlaValGlnIleMet 660
QY 1867 GTACTTAGACAAACAGTACCAA 1887
Db 661 ValLeuArgGlnTyrGln 667
```

Search completed: February 14, 2006, 17:19:42  
Job time : 288.633 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 14, 2006, 16:19:01 ; Search time 1.67839 Seconds  
(without alignments)  
2992.933 Million cell updates/sec

Title: US-10-723-552-3\_COPY\_5620\_7533

Perfect score: 3468

Sequence: 1 ATGATCCAGCTTAACCG.....CTAGCAGGAGAGCTGGCCGC 1914

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 194028

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abes/ABSSWEB spool/US10723552/runat 14022006 125151 13368/app query.fasta\_1  
-DB=Published Applications AA New -QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1  
-LOPCU=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs02p  
-USER=US10723552 @C9N 1.1 17 @runat 14022006 125151 13368 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -JONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DEPOP=6 -DEEXT=7

Database : Published Applications AA New:

- 1: /cgn2\_6/ptodata/1/pubpaa/US08 NEW PUB pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US06 NEW PUB pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US07 NEW PUB pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US05 NEW PUB pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10 NEW PUB pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11 NEW PUB pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US10 NEW PUB pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US06 NEW PUB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1267.5	36.5	654	US-10-510-947-6	Sequence 6, Appli
2	413.5	11.9	540	US-10-821-234-1395	Sequence 1395, Ap
3	250	7.2	229	US-11-014-842A-35	Sequence 1, Appli
4	170.5	4.9	1189	US-10-821-234-1209	Sequence 1209, Ap
5	127.5	3.7	676	US-10-510-947-1	Sequence 1, Appli
6	118	3.4	490	US-10-131-826A-310	Sequence 310, App
7	101.5	2.9	1225	US-11-102-476-2	Sequence 2, Appli
8	100	2.9	566	US-11-033-039-1244	Sequence 1244, Ap
9	97.5	2.8	748	US-11-098-686-10863	Sequence 10863, A

10	97	2.8	410	6	US-10-821-234-1180	Sequence 1180, Ap
11	96.5	2.8	801	7	US-11-014-842A-35	Sequence 35, Appli
12	96.5	2.8	869	7	US-11-014-842A-27	Sequence 27, Appli
13	96	2.8	742	7	US-11-175-690-528	Sequence 528, App
14	96	2.8	742	7	US-11-175-690-534	Sequence 534, App
15	96	2.8	742	7	US-11-175-690-546	Sequence 546, App
16	95	2.7	1985	6	US-10-495-083-7	Sequence 7, Appli
17	93.5	2.7	1255	7	US-11-022-562-235	Sequence 235, App
18	93.5	2.7	1255	7	US-11-052-554A-265	Sequence 265, App
19	93.5	2.7	1255	7	US-11-052-554A-266	Sequence 266, App
20	93.5	2.7	1255	7	US-11-052-554A-267	Sequence 267, App
21	93.5	2.7	1255	7	US-11-052-554A-268	Sequence 268, App
22	93.5	2.7	1255	7	US-11-052-554A-269	Sequence 269, App
23	93.5	2.7	1255	7	US-11-052-554A-270	Sequence 270, App
24	93.5	2.7	1255	7	US-11-052-554A-271	Sequence 271, App
25	93.5	2.7	1255	7	US-11-052-554A-272	Sequence 272, App
26	93.5	2.7	1255	7	US-11-052-554A-274	Sequence 274, App
27	93.5	2.7	1255	7	US-11-052-554A-275	Sequence 275, App
28	93.5	2.7	1279	6	US-10-957-880-3	Sequence 3, Appli
29	92.5	2.7	595	7	US-11-102-476-33	Sequence 33, Appli
30	92.5	2.7	3011	6	US-10-985-205-3	Sequence 3, Appli
31	92.5	2.7	8746	7	US-11-098-686-10232	Sequence 10232, A
32	92	2.7	515	7	US-11-052-554A-255	Sequence 255, App
33	92	2.7	1237	7	US-11-052-554A-95	Sequence 95, Appli
34	92	2.7	3194	7	US-11-052-554A-90	Sequence 90, Appli
35	91.5	2.6	922	7	US-11-115-086-9	Sequence 9, Appli
36	90.5	2.6	398	6	US-10-517-939-210	Sequence 210, App
37	90.5	2.6	757	7	US-11-052-554A-378	Sequence 378, App
38	90.5	2.6	899	6	US-10-453-372-528	Sequence 528, App
39	90.5	2.6	5935	6	US-10-995-561-776	Sequence 776, App
40	90	2.6	567	7	US-11-033-039-1279	Sequence 1279, Ap
41	89.5	2.6	536	7	US-11-076-074-2	Sequence 2, Appli
42	89.5	2.6	898	6	US-10-453-372-522	Sequence 522, App
43	89.5	2.6	899	6	US-10-453-372-506	Sequence 506, App
44	89.5	2.6	899	6	US-10-453-372-524	Sequence 524, App
45	89.5	2.6	899	6	US-10-453-372-526	Sequence 526, App

ALIGNMENTS

RESULT 1  
US-10-510-947-6  
; Sequence 6, Application US/10510947  
; Publication No. US2005055123A1  
; GENERAL INFORMATION:  
; APPLICANT: The Trustees of the University of Pennsylvania  
; APPLICANT: Wilson, James M.  
; APPLICANT: Medina, Maria Fe C.  
; APPLICANT: Kobinger, Gary  
; TITLE OF INVENTION: Chimeric Ebola Virus Envelopes and Uses Therefor  
; FILE REFERENCES: UPN-02811PCT  
; CURRENT APPLICATION NUMBER: US/10/510,947  
; CURRENT FILING DATE: 2004-10-28  
; PRIOR APPLICATION NUMBER: US 60/376,480  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/385,704  
; PRIOR FILING DATE: 2002-06-04  
; PRIOR APPLICATION NUMBER: US 60/427,752  
; PRIOR FILING DATE: 2002-11-20  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6  
; LENGTH: 654  
; TYPE: PRT  
; ORGANISM: Murine leukemia virus  
US-10-510-947-6  
Alignment Scores:  
Pred. No.: 4.15e-111 Length: 654  
Score: 1267.50 Matches: 277  
Percent Similarity: 57.0% Conservatives: 91  
Best Local Similarity: 42.9% Mismatches: 189  
Query Match: 36.5% Indels: 89





233 HisThrSerAsnLeuThrCysVal-----LysPheSerAsnThrThrTyr 247  
1045 TCTGGAAAGGCGACCTGCATAGGAAGAAGTTCCCCCATCCCAACAACACCTTTGTGAACCAC 1104  
::: ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
248 ThrAsnSerGlnCysIleArgTrpValThrProThrGlnIleValCysLeuPro 267  
1105 ACTGAAGCCCTTTAATCAAACCTCTGAGAGTCATATCTGGTACTCTGGTTATGACAGGTGG 1164  
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
268 SerGlyIlePhe----- 271  
1165 TGGGCATGTAATACTAGATTAAACCCCTTGTTGGTTTCCACCTTGGTTTTAACCAAACTAAA 1224  
Phe:: |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
272 PheValCysGlyThrSerAlaIleArgCysLeuAsn-----GlySerSerGlu 287  
1225 GATTTTTCATTATGTCCAAATGTTTCCCGAGTGTATTACTATCCGAAAAAGCAATC 1284  
||| ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
288 SerMetCysPheLeuSerPheLeuValProMetThrIleThrGluGln----- 305  
1285 CTTGATGAATGACTAC-----AGAAATCATCGACAAAGAGAGAACCCATATCTCTG 1338  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
306 ---AspLeuTyrSerTyrValIleSerLysProArgAsnLysArgValProIle----- 322  
1339 ACATCTGCTGTGATGCTCGGACTTGGAGTGGCAGAGGTGTAGGAACAGAGAACAGCTGCC 1398  
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
323 ---LeuProPheValIleGlyAlaGlyValLeuGlyAlaLeuGlyThrGlyIleGlyGly 341  
1399 CTGCTCACGGGACCACAGCAGCTAGAAACAGGACTTAGTAACCTACATCGAATTGTAACA 1458  
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
342 IleThrThrSerThrGln-----PheTyrTyrLysLeuSerGlnGluLeuAsn 357  
1459 GAAGATCTCCAAGCCCTAGAAAAATCTGCAGTAACCTCGAGGAATCCCTAAACCTCCTTA 1518  
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
358 GlyAspMetGluArgValAlaAspSerLeuValThrLeuGlnAspGlnLeuAsnSerLeu 377  
1519 TCTGAAGTAGTCTCAGCAATAGAACAGGGTGTAGATTATTATTTCTTAAAGAGGAGGA 1578  
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
378 AlaAlaValValLeuGlnAsnArgArgAlaLeuAspLeuThrAlaGluArgGlyGly 397  
1579 TTATGTGTAGCTTGAAGAGGAATCTGTTTATTGTGATCATTCAGGGGCATCAGA 1638  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
398 ThrCysLeuPheLeuGlyGluGluCysCysTyrTyrValAsnGlnSerGlyIleValThr 417  
1639 GACTCATGAACAAGCTTAGAGAAAGTTGGAGAACGCTGCAAGGGGAAAAAGGAACACTACT 1698  
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
418 GluLysValLysGluIleArgAspArgIleGlnArgArgAlaGluGluLeuArgAsnThr 437  
1699 CAAGGTGGTTGAGGGATGGTTCAACAGGTCTCTTGGTTGGCTACCCCTACTTTCTGCT 1758  
438 GlyProTrp-----GlyLeuLeu-----SerGlnTrpMetProTrpIleLeuProPhe 453  
1759 TTAACAGGACCTTAATAGTCTCCTCCGTGTACTACAGTTGGGCGCATGATTATTAAAC 1818  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
454 Leu---GlyProLeuAlaIleIleLeuLeuLeuPheGlyProCysIlePheAsn 472  
1819 AAGTTAAATTCCTTCAATAGAGAACGAATAAGTGCAGTCCAGTC---ATGGTACTTTAGA 1875  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
473 LeuLeuValAsnPheValSerArgIleGluAlaValLysLeuGlnMetGluProLys 492  
1876 CAACAG-----TACCNAAGCCGCTAGCAGGGAAGCT 1908  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
493 MetGlnSerLysThrLysIleTyrArgProLeuAspArgProAla 508

RESULT 3  
US-11-014-629-1  
; Sequence 1, Application US/11014629  
; Publication No. US20050244376A1  
; GENERAL INFORMATION:  
; APPLICANT: Hall, Frederick L.  
; APPLICANT: Gordon, Erlinda M.  
; TITLE OF INVENTION: TARGETED GENE DELIVERY TO EXTRACELLULAR MATRIX  
; FILE REFERENCE: 30863-704.302  
; CURRENT APPLICATION NUMBER: US/11/014,629  
; CURRENT FILING DATE: 2004-12-15

```

RESULT 3
US-11-014-629-1
; Sequence 1. Application US/11014629
; Publication No. US2005024376A1
; GENERAL INFORMATION:
; APPLICANT: Hall, Frederick L.
; APPLICANT: Gordon, Erlinda M.
; TITLE OF INVENTION: TARGETED GENE 1
; FILE REFERENCE: 30863-704.302
; CURRENT APPLICATION NUMBER: US/11/
; CURRENT FILING DATE: 2004-12-15

```

```

; PRIOR APPLICATION NUMBER: US 08/837,223
; PRIOR FILING DATE: 1997-04-10
; PRIOR APPLICATION NUMBER: US 09/904,923
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 229
; TYPE: PRT
; ORGANISM: moloney murine leukemia virus
US-11-014-629-1

```

Alignment Scores:	
Pred. No.:	1,05e-15
Score:	250.00
Percent Similarity:	39.1%
Best Local Similarity:	29.4%
Query Match:	7.2%
DB:	7
Length:	229
Matches:	70
Conservative:	23
Mismatches:	75
Indels:	12
Gaps:	1

US-10-723-552-3 COPY 5620 7533 (1-1914) x US-11-014-629-1 (1-229)

157	QY	CATAAACCCTTATCTCTCACCTGGTTAAATTA	CTGACTCCGGCACA	CAGGTATTATATCAAC	216
8	Db	HisGlnValTyrAsnIleThrTrpGluValThr	--AsnGlyAspArgGluThrValTrp	26	
217	QY	AACACTCAAGGGAGGCGCTCCTTTAGGAACCTGGTGGCTGATCTATAC	-----GTTTGC	270	
27	Db	AlaThrSerGlyAsnHisProLeuThrTrpTrpProAspLeuThrProAspLeuCys	46		
271	QY	CTC-----AGATCAGTTATTTCCCTAGT	291		
47	Db	MetLeuAlaHisGlyProSerTyrTrpGlyLeuGluTyrGlnSerProPheSerSer	66		
292	QY	CTGACCTCACCCCCA-----	306		
67	Db	ProProGlyProProCysCysSerGlySerSerProGlyCysSerArgAspCysGlu	86		
306	QY	-----	306		
87	Db	GluProLeuThrSerLeuThrProArgCysAsnThrAlaTrpAsnArgLeuLysLeuAsp	106		
307	QY	GATATCTCTCATGCT-----CACGGATTATTATTTGCCCAGGACCA-----CCAAAT	354		
107	Db	GlnThrThrHisLysSerAsnGluGlyPheTyrValCysProGlyProHisAspProArg	126		
355	QY	AATGGAAAAATTGGCGAAATCCACAGAGATTCTTTGTAAACAATCGAACTGTGAAC	414		
127	Db	GluserLysSerCysGlyGlyProAspSerPheTyrCysAlaTyrTrpGlyCysGluThr	146		
415	QY	TCTAATGATGGATATTGGAAATGGCCACCTCTCAGCAGGATAGGGTAAGTTTCTTAT	474		
147	Db	ThrGlyArgAlaTyrTriplys---ProSerSerTrpAsp-----PheIleThr	162		
475	QY	GTCAACACCTATACGAGCTCTGGCAAAATTAATACCTGACCTGGATTAGAACTGGAAGC	534		
163	Db	ValAsnAsnLeuThrSerAspGln-----AlaValGlnValCysLysAspAsn	179		
535	QY	CCCAAGTGTCTCTCTTCAGACCTAGATTACCTAAAAATAAGTTTCACTGAGAAGGAAA	594		
180	Db	LysTrpCysAsnPro-----LeuValIleArgPheThrAspAlaGlyArg	194		
595	QY	CAAGAAAATATCCTAATATGGGTAATGGTATGCTTGGGGAATGGTATTATTATGAGGC	654		
195	Db	Arg-----ValThrSerTrpThrGlyHisTyrTrpGlyLeuArgLeuTyr---Val	211		
655	QY	TGGGGTAAACAAACGAGCTCCCATCTTAACATTTCGCTCAAAATAAACACGAGCTG	708		
212	Db	SerGlyGlnAspProGlyLeuThrPheGlyIleArgLeuArgTyrGlnAsnLeu	229		

RESULT 4

US-10-821-234-1209

Sequence 1209. Application US/10821234

```

; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1209
; LENGTH: 1189
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1189)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-821-234-1209

Alignment Scores:
Pred. No.: 4-92e-08 Length: 1189
Score: 170.50 Matches: 45
Percent Similarity: 48.1% Conservative: 17
Best Local Similarity: 34.9% Mismatches: 34
Query Match: 4.9% Indels: 33
DB: 6 Gaps: 4

```

US-10-723-552-3 COPY 5620 7533 (1-1914) x US-10-821-234-1209 (1-1189)

Qy	1468	CAAGCCCTAGAAAA-----TCGTGCTAGTAACCTGGAGGAATCCCTAACCC-----	1512
Db	1075	GlnAlaLeuProLysProHisGluGlyValSerAsnProLys**ThrCysSerProIle	1094
Qy	1513	-----TCCCTTACTGAAGTAGTCTACAGAAATAGAGAGGGTTAGATTATTATTT	1563
Db	1095	ProLysAspSerLeuSerArgValThrLeuGlnAsnHisArgGlyLeuAspLeuLeuThr	1114
Qy	1564	CTAAAGCAAGGAGGATTATGTGAGCCTTCAAGAGGAAATGCTCTTTTATGTGGATCAT	1623
Db	1115	AlaGluLysGlyGlyLeuCysIlePheLeuGluGluGluCysCysPheTyrThrAsnGln	1134
Qy	1624	TCAGGGCCCATCAGAGACTCCATGAACAACGTATTAGAAAAGTTGGAGAGCGTCGAAGG	1693
Db	1135	SerGlyLeuValGlnAspAlaAlaGlyArgIleAsnGluLysAlaSer-----	1150
Qy	1684	GAAGAAGAACTACTCAAGGCTGTTTCAGGGATGGTTCACAGGTCTCTTTGGTTGGCT	1743
Db	1151	-----GlyArgValGlnTrpLeuThr	1157
Qy	1744	ACCTACTTTCTGCTTTA-----ACAGACCCCTTAATAGTC	1779
Db	1158	ProValIleProaspLeuTrpGluAlaGluAlaGlySerArgGlyGlnGluIleGlu	1177
Qy	1780	CTCCTCTCTGTACTCACAGTTGGGCCA	1806
Db	1178	ThrIleLeuAlaAsnThrValLysPro	1186
RESULT 5			
US-10-510-947-1			
; Sequence 1, Application US/10510947			
; Publication No. US20050255123A1			
; GENERAL INFORMATION:			
; APPLICANT: The Trustees of the University of Pennsylvania			
; APPLICANT: Wilson, James M.			
; APPLICANT: Medina, Maria Fe C.			
; APPLICANT: Kobinger, Gary			
; TITLE OF INVENTION: Chimeric Ebola Virus Envelopes and Uses Therefor			
; FILE REFERENCE: UPN-02811PCT			

;; CURRENT APPLICATION NUMBER: US/10/510,947  
;; CURRENT FILING DATE: 2004-10-28  
;; PRIOR APPLICATION NUMBER: US 60/376,480  
;; PRIOR FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/385,704  
;; PRIOR FILING DATE: 2002-06-04  
;; PRIOR APPLICATION NUMBER: US 60/427,752  
;; PRIOR FILING DATE: 2002-11-20  
;; NUMBER OF SEQ ID NOS: 67  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 1  
;; LENGTH: 676  
;; TYPE: PRT  
;; ORGANISM: Ebola virus  
US-10-510-947-1

Alignment Scores:  
Pred. No.: 0.000455 Length: 676  
Score: 127.50 Matches: 103  
Percent Similarity: 36.1% Conservative: 88  
Best Local Similarity: 19.5% Mismatches: 206  
Query Match: 3.7% Indels: 132  
DB: 6 Gaps: 23

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-10-510-947-1 (1-676)

```
Qy 439 CCAACCTCTCAGCAGATAGGGTAAAGTTTCTTATGTCAACACC-----TATACC 489
Db 202 ProValAsnAlaThrGluAspProSerSerGlyTyrThrThrThrIleArgTyrGln 221
Qy 490 AGCTCTCGA-----CAATTAATACCTGACCTGG 519
Db 222 AlaThrGlyPheGlyThrAsnGluThrGluTyrLeuPheGluValAspAsnLeuThrTyr 241
Qy 520 ATTAGAAGTGGAGCCCAAGTGTCTCTCAGACCTAGATTACCTAAATAAAGTTTC 579
Db 242 ValGlnLeuGluSer---ArgPheThrProGlnPheLeuGlnLeuAsnGluThrIle 260
Qy 580 ACTGAGAAAGGAAACCAAGAAATATC-----CTAAATGG----- 615
Db 261 TyrThrSerGlyLysArgSerAsnThrThrGlyLysLeuIleTyrLysValAsnProGlu 280
Qy 616 GTAATGGTATGCTTGGGAATGGTATATATAGAGGCTCGGGTAACCAACACAGGCTCC 675
Db 281 IleAsp---ThrThrIleGlyGluTrp-AlaPheTyrGlu-----ThrLysLysA 296
Qy 676 ATTCTAACTATTCGCTCAAAATAAACACAGCTGGAGCTCCAATGGCTATAGACCAAT 735
Db 296 sn-LeuThrArgLysIleArgSerGluLeuSerPheThrValValSerAsnGlyAla 315
Qy 736 ACGGTCTTGACGGGTCAAAGACCCCAACCAAGGACGACCATCTCTAACTAACT 795
Db 316 LysAsnIleSerGlyGlnSerProAlaArgThrSerSerAspProGlyThrAsnThrThr 335
Qy 796 TCTGGATCAGACCCCACTGAGTCTAACACGACGACTAAATAGGGGGCAAACTTTTTC 855
Db 336 ThrGluAspHisLysIleMetAlaSerGluAsnSerSerAlaMetValGlnValHisSer 355
Qy 856 CTATCCAGGGAGCTTTTCAAGCTCTTAACCTCACAGCTCCAGGAGCTACTCTCTTGT 915
Db 356 -----GlnGlyArgGluAlaAlaValSerHisLeuThrThrLeuAlaThr----- 370
Qy 916 TGGCTATGCTTAGTCTCGGGCCCACTTACTATGAGGAATGGCTAGAGGGAATTC, 975
Db 371 -----IleSerThrSerPro-----GlnSer 377
Qy 976 AATGTGCAAAAGAACATAGACCAATGCACATGGGGATCCCAA---AATAAGCTTACC 1032
Db 378 LeuThrThrLysProGlyProAspAsnSerThrHisAsnThrProValTyrLysLeuAsp 397
Qy 1033 CTTACTAGAGTTTCTTGAAAGGACCTGCATAGGAAGGTTCCCATCCCAACACAC 1092
Db 398 IleSerGluAlaThr-----GlnValGluGlnHisArgArg 410
```

```
Qy 1093 CTTTGTAAACCACTGAAGCCTTTAATCAAACTCTGAGAGTCAATATCTGTACTCGT 1152
Db 411 ThrAspAsnAspSerThrAlaSerAspThrProSerAlaThrThrAlaAlaGlyPro--- 429
Qy 1153 TATCAGAGGTGGTGGGCATGTAATACTGGATTAAACCTTGTGTGTTCACCTGGTTT 1212
Db 430 -----ProLysAlaGluAsnThrAsnThr 437
Qy 1213 AACCAAAATAAGATTTTTGCATTATGTCTCCAAATTGTTCCCGAGTGTATTACTATCC 1272
Db 438 SerLysSerThrAspPhe-----LeuAspProAlaThrThrThrSerPro 452
Qy 1273 GAAAAACAATCTCTTGATGAATATGACTACAGAAATCATCGACAA----- 1317
Db 453 GlnAsnHisSerGluThrAlaGlyAsnAsnAsnThrHisHisGlnAspThrGlyGluGlu 472
Qy 1318 -----AAGAGAGAACCCTATCTCTGACACTTGTGTGTGTCTGGACTT 1362
Db 473 SerAlaSerSerGlyLysLeuGlyLeuIleThrAsnThrIleAlaGlyValAlaGlyLeu 492
Qy 1362 ----- 1362
Db 493 IleThrGlyArgArgThrArgArgGluAlaIleValAsnAlaGlnProLysCysAsn 512
Qy 1363 -----GGAGTGGCAGCAGGTGTAGGA----- 1383
Db 513 ProAsnLeuHisTyrTrpThrThrGlnAspGluGlyAlaIleGlyLeuAlaTrpIle 532
Qy 1384 -----ACAGAAACAGCTGCC-----CTGTCTACGGGACACACAGCAGCTAGAAACA 1428
Db 533 ProTyrPheGlyProAlaAlaGluGlyIleTyrIleGluGlyLeuMetHisAsnGlnAsp 552
Qy 1429 GGACTT---AGTAACCTACATCGAATTGTAAACAGAGATCTCCAGCCCTAGAAAAATCT 1485
Db 553 GlyLeuIleCysGlyLeuArgGlnLeuAlaAsnGluThrThrGlnAlaLeuGlnLeuPhe 572
Qy 1486 GTCAAGTAACCTGGAGGAATCCCTAACTCTTCTGAAGTAGTCTTACAGATAGAGA 1545
Db 573 Leu-----ArgAlaThrThrGluLeuArgThrPheSerIleLeuAsnArgLys 588
Qy 1546 GGGTTAGATTATTTCTTAAAGAGAGGAGGATTATGTGTAGCTTTGAAAGGAGGATGC 1605
Db 589 AlaIleAspPheLeuLeuGlnArgTyrGlyThrCysHisIleLeuGlyProAspCys 608
Qy 1606 TGTTTTATCTG---GATCATTGAGGGCCATCAGACACTCCATGAACAAG---CTTAGA 1659
Db 609 CysIleGluProHisAspTrpThrLysAsnIleThrAspLysIleAspGlnIleHis 628
Qy 1660 GAAAGGTGGAGAGCGTCCGAAAGGAAAGAACTACTCAAGGGTGGTTTGGGGATGG 1719
Db 629 AspPheValAspLysThrLeuProAspGlnGlyAspAsnAspAsnTrpThrGlyTrp 648
Qy 1720 TTCAACAGGTCTCTTTGGTGGTACCCCTACTTTCTGCTTTTAAACAGACCCCTTAATAGTC 1779
Db 649 -----ArgGlnTrpIleProAlaGlyIle---GlyValThrGlyValIleAla 664
Qy 1780 CTCCTCTCTTACTACAGTT 1800
Db 665 ValIleAlaLeuPheCysIle 671
```

## RESULT 6

US-10-131-826A-310  
; Sequence 310, Application US/10131826A  
; Publication No. US20050245730A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.





Db 213 Gln-----AlaSerGly----- 216  
Qy 538 AAGTGTCTCTTCAGACCTAGATTACCTAAAATAAGTTTCACTGAGAAAGGAAACAA 597  
Db 217 ::::::::::: 597  
Qy 598 GAAATATATCTTA-----ArgValThrValSerThrArgSerGln 226  
Db 227 GlnThrIleIleProAsnIleGlySerArgProTrpValArgGlyLeuSerSerArgIle 246  
Qy 631 -----TGGGAATGGTATATTATGAGAGCTCGGGTAAACACACAGCTCCATCTTA 681  
Db 247 SerIleTyrTrpThrIleVal-----LysProGlyAspValLeu 259  
Qy 682 ACTATTGCCTCAAAATAAACACGCTGAGCCT-----CCAAATGGCTATAGGA 729  
Db 260 ValIleAsnSerAsnGlyAsnLeuIleAlaProArgGlyTyrPheLysMetArgThrGly 279  
Qy 730 CCAATACGGTCTTACCGGGTCAAGAGACCCCAACCAAGCAGGACCACTCTCTAAC 789  
Db 280 LysSerSerIleMetArgSerAspAlaPro-----IleAspThrCysIleSerGluCys 297  
Qy 790 ATAACCT---TCTGATCAGACCCCACTGAGTCTAACACGACGACTAAATGGGGCAAAA 846  
Db 298 IleThrProAsnGlySerIleProAsnAspLys----- 308  
Qy 847 CTTTTCAGCCTCATCCAGGAGCTTTTCAAGCTCTTAACCTCCACGACTCCAGAGCTACC 906  
Db 309 :::::::::::ProPheGlnAsnValAsnLysIleThrGlyAla--- 320  
Qy 907 TCTTCTTGTGGCTATGTAGCTTCGGGCCCACTTACTATGAAGGAATGGCTAGAAGA 966  
Db 321 -----Cys-----ProLysTyrValLys----- 326  
Qy 967 GGGAAATTCAATGTGACAAAAGAACATAGAGACCAATGCATGGGGATCCCAAAATAAG 1026  
Db 327 -----GlnAsnThr 329  
Qy 1027 CTTACCCCTTACTGAGTTTCTGGAAAAGCGACCTGCATAGGAAGGTTCCCCCATCCAC 1086  
Db 330 LeuLysLeu-----AlaThrGlyMetArgAsnValPro----- 340  
Qy 1087 CAACACCTTTGTAACCACTGAAGCCTTTAATCAAACTCTGAGAGTCAATATCTGTA 1146  
Db 341 -----GluLysGln----- 343  
Qy 1147 CTGTTTATGACAGGTGGTGGCATGTAATACTGGAATTAACCCCTTGTTGTTCCACCTTG 1206  
Db 344 -----ThrArgGlyLeuPheGlyAlaIleAlaGlyPhe 354  
Qy 1207 GTTTTAAACCAACTAAAGATTTTGGCAATTATGGTCCAAATGTTCCCGAGTGATTATAC 1266  
Db 355 IleGluAsnGlyTrp----- 359  
Qy 1267 TATCCGAAAAGCAATCTTGATGAATGACTACAGAAATCATCGAATAAGAGAGAA 1326  
Db 360 -----GluGlyMetIleAspGlyTrpTyrGlyPheArgHisGlnAsnSerGlu----- 375  
Qy 1327 CCCATATCTCTGACACTTGTCTGTGATGCTCGGACTTGGAGTGGCAGCAGGTGTAGGAACA 1386  
Db 376 -----GlyThrGlyGlnAlaAlaAspLeuLysSer 385  
Qy 1387 GGAACAGCTGCCCTGGTCACGGGACCAAGCAGCTAGAAACAGGACTTAGTAACCTACAT 1446  
Db 386 ThrGlnAlaAlaIle-----AspGlnIleAsnGlyLysLeuAsnArgValIle 401  
Qy 1447 CGAATTGTAACAGAGATCTCCAAGCCCTAGAAAATCTGTCAGTAACCTCGAGGAATCC 1506  
Db 402 GluLysThrAsnGluLysPheHisGlnIleGluLysGluPheSerGluValGluGlyArg 421  
Qy 1507 CTAACCTCCTTATCTGAAGTAGTCTCAGCAATAGAGAGGGTTAGATTATTATTCTTA 1566  
Db 422 IleGlnAspLeuGluLysTyrValGluAspThrLys-----IleAsp---LeuTrpSer 438

Qy 1567 AAAGAAGGAGGATTATGTAGCTTGAAGGAGGAATCTGCTTTTATGTGGATCATCA 1626  
Db 439 TyrAsnAlaGluLeuLeuValAlaLeuGluAsnGln-----HisThr 452  
Qy 1627 GGGCCCATCAGAGACTCC---ATGAACAAGCTTAGAAGAAAGTTGGAGAACGCTCGAAGG 1683  
Db 453 IleAspLeuThrAspSerGluMetAsnLysLeuPheGluLysThrArgGlnLeuArg 472  
Qy 1684 GAAAGGAAACTACTCAAGGG 1704  
Db 473 GluAsnAlaGluGluMetGly 479

## RESULT 9

US-11-098-686-10863  
; Sequence 10863, Application US/11098686  
; Publication No. US20060024696A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES  
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING  
; FILE REFERENCE: 09531-128001  
; CURRENT APPLICATION NUMBER: US/11/098,686  
; PRIOR FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: PCT/US03/31318  
; PRIOR FILING DATE: 2003-10-01  
; PRIOR APPLICATION NUMBER: US 60/416,395  
; PRIOR FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 11433  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10863  
; LENGTH: 748  
; TYPE: PRT  
; ORGANISM: Lawsonia intracellularis  
US-11-098-686-10863

Alignment Scores:  
Pred. No.: 0.307 Length: 748  
Score: 97.50 Matches: 88  
Percent Similarity: 34.5% Conservative: 54  
Best Local Similarity: 21.4% Mismatches: 140  
Query Match: 2.8% Indels: 129  
Gaps: 19

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-11-098-686-10863 (1-748)

Qy 511 CTGACCTGGATTAGAACTGGAAGCCCAAGTGCTCTCTCAGACCTAGATTACCTAAA 570  
Db 135 LeuGluPheIleSerTyrGlyLeuProHisSerGluLeuGlyAsnAspPheValLys 154  
Qy 571 ATAAGTTTCACCTGAGAAGGAAACAAAGAAATATCCTA----- 609  
Db 155 ---LysAlaGluLysSerSerGluLysAsnProLeuThrLysTyrThrIleGluLeu 173  
Qy 610 -----AAATGGTAAATGGTATGTTGGGGAATGTTATATATGGAGGCTCGGGTAAA 663  
Db 174 ThrGlnLysAlaArgAsnGlyGluIleAspProLeuValGlyArgAspMetGluIleSer 193  
Qy 664 CAACAGGCTCCATTCTAACTATTTCGCCTCAAAATAAACACGAGCTGGAGCCTCCATGGCT 723  
Db 194 ArgThrIleGluValLeuAlaArgLysLysAsnAsn-----ProLeuTyr 209  
Qy 724 ATAGA---CCAAATACGGTCTTCAGCGGTCAAGAGACCCCAACCAAGGACCAAGCA 780  
Db 210 ValGlyAspProGlyThrGlyLysThr-----AlaIleAlaGluGlyLeu 224  
Qy 781 TCCTCTAACATACTTCTGGATCAGACCCCACTGAGTCTAACAGCAGCACTAAATATGGG 840  
Db 225 AlaLeuArgIleIleSerGlyAsnValProGluGluPheThrSerSerLeu----- 241  
Qy 841 GCAAAACCTTTTAGCTCATCTCAGGAGCTTTTCAAGCTCTTAAGCTCTTCAAGCTCCAGAG 900  
Db 242 -----IlePheSerLeuAspMetGlyAla----- 249







```
Db 603 PheAspIleThrAsnTrpLeuTrpTyrIleArgLeuPheIleMetIleValGlyGlyLeu 622
Qy 1774 ATAGTCCTCTCTCTACTCACAGTTCGGCCATGATTATTATAACAAGTTAATTGCGCTTC 1833
Db 623 IleGlyLeuArgIleValPheAlaVal---LeuSerIleValAsnArg----- 637
Qy 1834 ATTAGAGAACGAATAAGTCAGTCAGATCATGCTACTTACAGCAACAGTACCAAGCCG 1893
Db 638 ValArgGlnGlyTyrSerProLeuSerPheGlnThrLeuLeuProAlaSerArgGlyPro 657
Qy 1894 TCTAGCAGGGAA 1905
Db 658 AspArgProGlu 661

RESULT 12
US-11-014-842A-27
; Sequence 27, Application US/11014842A
; Publication No. US20060013826A1
; GENERAL INFORMATION:
; APPLICANT: TANGY, FREDERIC
; APPLICANT: LORIN, CLARISSE
; APPLICANT: MOLLET, LUCILE
; APPLICANT: DELEBEQUE, FREDERIC
; TITLE OF INVENTION: RECOMBINANT MEASLES VIRUSES EXPRESSING EPITOPES OF
; TITLE OF INVENTION: ANTIGENS OF RNA VIRUSES - USE FOR THE PREPARATION OF
; TITLE OF INVENTION: VACCINE COMPOSITIONS
; FILE REFERENCE: 2356.0093
; CURRENT APPLICATION NUMBER: US/11/014,842A
; CURRENT FILING DATE: 2004-12-20
; PRIOR APPLICATION NUMBER: PCT/EP03/07146
; PRIOR FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: EP 02291550.8
; PRIOR FILING DATE: 2002-06-20
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn ver. 3.3
; SEQ ID NO 27
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-11-014-842A-27

Alignment Scores:
Pred. No.: 0.399 Length: 869
Score: 96.50 Matches: 100
Percent Similarity: 30.9% Conservative: 62
Best Local Similarity: 19.1% Mismatches: 185
Query Match: 2.8% Indels: 177
DB: 7 Gaps: 22

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-11-014-842A-27 (1-869)
Qy 769 GGACGAGACCATCTCTAACAATA-----ACTTCTGATCAGCCCACT 813
Db 238 GlySerGlyProCysThrAsnValSerThrValGlnCysThrHisGlyIleArgProVal 257
Qy 814 GAGTCTAACACGACGACTAAATCGGGSCA----- 843
Db 258 ValSerThrGlnLeuLeuLeuAsnGlySerLeuAlaGluLeuAspIleValIleArgSer 277
Qy 844 -----AAACTTTTTCAGCTCATCCAGGAGGCTTTTCAAGCTCTT 882
Db 278 GluAspPheThrAspAsnValLysThrIleIleValGlnLeuAsnGluSerValValIle 297
Qy 883 AACTCCACGACTCCAGAGGCTACTCTCTTCTTGTGCTGCTAGCTTCGGGCCCACT 942
Db 298 AsnCysThrArgProAsnAsnThrArgGluArgLeuSerIleGlyProGlyArgAla 317
Qy 943 TACTATGAA-----GGAATGGCTAGAGGGAATTCATGTGCACAAA 987
Db 318 PheTyrAlaArgArgAsnIleIleGlyAspIleArgGlnAlaHisCysAsnIleSerArg 337
Qy 988 GAACATAGAGACCAATGCATGGGATCCCAAAATAAGCTTACCCTTACTGAGTTTCT 1047
```

```
Db 338 -----AlaLysTrpAsnAsn-----ThrLeuGlnGlnIle--- 347
Qy 1048 GGAAGAGCAGCCTCATAGGAAGGTTCCCCCTCCACCAACACCTTTTGTAAACACACT 1107
Db 348 -----ValIleLysLeuArgGluLysPheArgAsnLysThr 359
Qy 1108 GAAGCCTTTTAATCAAACTCT-----GAGAGTCAATATCTGGTACCTGGTTATGACAGG 1161
Db 360 IleAlaPheAsnGlnSerSerGlyAspProGluIleValMetHisSerPheAsn--- 378
Qy 1162 TGTGGGCGATGTAATACTGGATTAAACCCCTTGTGTTCCACCTTGGTTTAAACCAACT 1221
Db 379 -----CysGlyGlyGluPhePheTyrCysAsnThrAlaGlnLeuPheAsnSerThr 395
Qy 1221 ----- 1221
Db 396 TrpAsnValAlaGlyGlyThrAsnGlyThrGluGlyAsnAspIleIleThrLeuGlnCys 415
Qy 1222 -----AAAGATTTTTCATTTATGTCCTCAAAATTTCTCCCGAGGTGTTATCTATCCC--- 1272
Db 416 ArgIleLysGlnIleIleAsnMetTrpGlnLysValGlyLysAlaMetTyrAlaProPro 435
Qy 1272 ----- 1272
Db 436 IleThrGlyGlnIleArgCysSerSerAsnIleThrGlyLeuLeuLeuThrArgAspGly 455
Qy 1273 -----GAAAAGCAATCCTT-----GATGAATATGAC 1299
Db 456 GlyAsnSerThrGluThrGluThrGluIlePheArgProGlyGlyGlyAspMetArgAsp 475
Qy 1300 -----TACAGAAATCATCGACAAAGAGAGAACCATATCTCTGACA 1341
Db 476 AsnTrpArgSerGluLeuTyrLysValValArgIleGluProIleGlyValAla 495
Qy 1342 CTT-----GCT 1347
Db 496 ProThrArgAlaLysArgArgThrValGlnArgGluLysArgAlaValGlyIleGlyAla 515
Qy 1348 GTGATGCTCGGACTTGGAGTGGCAGCAGGTGTAGGAACAGACAGCTGCC----- 1398
Db 516 ValPheLeuGlyPheLeuGlyAlaAlaGlySerThrMetGlyAlaAlaSerValThrLeu 535
Qy 1399 -----CTGGTCACGGGACACAGCAGCTAGAAACAGGACTT----- 1434
Db 536 ThrValGlnAlaArgLeuLeuLeuSerGlyIleValGlnGlnGlnAsnAsnLeuLeuArg 555
Qy 1435 -----AGTAACTACATCGAATTGTAAACAGAGAATCTCCCAAGCCCTAGAA 1479
Db 556 AlaIleGluAlaGlnGlnAsnMetLeuArgLeuThrValTrpGlyIleLysGlnLeuGln 575
Qy 1480 AAATCTGTCTAGTAACCTGGAGGATCCCTA----- 1509
Db 576 AlaArgValLeuAlaLeuGluArgTyrLeuArgAspGlnGlnLeuMetGlyIleTrpGly 595
Qy 1510 -----ACCTCTCTATCTGAAGTAGCTTACAGAAATAGAGA 1545
Db 596 CysSerGlyLysLeuIleCysThrThrSerValProTrpAsnValSerTrpSerAsnLys 615
Qy 1546 GGGTTAGATTATTATTATTTCTAAAGAGGAGGATTATGTGTAGCTTGAAGAGGAGATGC 1605
Db 616 SerValAspAspIleTrp-----AsnAsnMetThrTrpMetGluTrpGluArgGluIle 633
Qy 1606 TGTTTTTATGTGATCATTCAGGGCCCATCAGAGACTCCATGAAACAGCTTAGAGAAAGG 1665
Db 634 AspAsnTyrThrAspTyr-----IleTyrAspLeuLeuGlyLysSerGlnThrGln 650
Qy 1666 TTGAGAGAGCGTCGAGGGGAAAAGAAACTACTCAAGGGTGGTTTGAG-----GGATGG 1719
Db 651 GlnGluLysAsnGluLysGluLeuLeuGluLeuAspLysTrpAlaSerLeuTrpAsnTrp 670
Qy 1720 TTCAACAGAGTCT-----CTTTGGTTGGCTACCTTCTCTGCTTTTAAACAGGACCCCTTA 1773
```

Db 671 PheAspIleThrAsnTrpLeuTrpTyrlleArgLeuPheIleMetIleValGlyGlyLeu 690  
Qy 1774 ATAGTCCTCCCTGTTACTCACAGTTCGGCCATGATTATTAAACAAGTTAATGCGCTTC 1833  
Db 691 IleGlyLeuArgIleValPheAlaVal---LeuSerIleValAsnArg- 705  
Qy 1834 ATTAGAGAACGAATAAGTCGATCCAGATCATGCTATTAGACACACATACCAAGCCG 1893  
Db 706 ValArgGlnGlyTySerProLeuSerPheGlnThrLeuLeuProAlaSerArgGlyPro 725  
Qy 1894 TCTAGCAGGAA 1905  
Db 726 AspArgProGlu 729

## RESULT 13

US-11-175-690-528  
; Sequence 528, Application US/11175690  
; Publication No. US20060014254A1  
; GENERAL INFORMATION:  
; APPLICANT: Haseltine et al.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF605  
; CURRENT APPLICATION NUMBER: US/11/175,690  
; CURRENT FILING DATE: 2005-07-07  
; PRIOR APPLICATION NUMBER: PCT/US04/001369  
; PRIOR FILING DATE: 2004-01-20  
; PRIOR APPLICATION NUMBER: US 60/441,305  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: US 60/453,201  
; PRIOR FILING DATE: 2003-03-11  
; PRIOR APPLICATION NUMBER: US 60/467,222  
; PRIOR FILING DATE: 2003-05-02  
; PRIOR APPLICATION NUMBER: US 60/472,816  
; PRIOR FILING DATE: 2003-05-23  
; PRIOR APPLICATION NUMBER: US 60/476,267  
; PRIOR FILING DATE: 2003-06-06  
; PRIOR APPLICATION NUMBER: US 60/505,172  
; PRIOR FILING DATE: 2003-09-24  
; PRIOR APPLICATION NUMBER: US 60/506,746  
; PRIOR FILING DATE: 2003-09-30  
; NUMBER OF SEQ ID NOS: 568  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 528  
; LENGTH: 742  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-11-175-690-528

## Alignment Scores:

Pred. No.:	0.424	Length:	742
Score:	96.00	Matches:	64
Percent Similarity:	32.0%	Conservative:	34
Best Local Similarity:	20.9%	Mismatches:	88
Query Match:	2.8%	Indels:	120
DB:	7	Gaps:	14

US-10-723-552-3\_COPY\_5620\_7533 (1-1914) x US-11-175-690-528 (1-742)

Qy 1030 ACCCTTACTAGGTTTCTGGAAGGCGACCTGCATAGGAAGGTTCCCCCATCCCAACAA 1089  
Db 446 ThrLeuValGluValSerArgAsn-----LeuGlyLysVal-----GlySer 459  
Qy 1090 CACCTTTGTACCACTGACAGCCCTTTAATCAACCTCTGAGATCAATATCTG----- 1143  
Db 460 LysCysCysLysHisProGluAlaLysArgMetProCysAlaGluAspTyLeuSerVal 479  
Qy 1144 -----GTACCTGGTTATGACAGGTGGGGCA 1170  
Db 480 ValLeuAsnGlnLeuCysValLeuHisGluLysThrProValSerAspArgValThrLys 499  
Qy 1171 TGTAATACT-----GGATTAACCCCTGTGTCTCCACCTGTGTTTAAACCAA 1218  
Db 500 CysCysThrGluSerLeuValAsnArgArgProCysPheSerAlaLeuGluValAspGlu 519

Qy 1219 ACT-----AAAGATTTTTGCATT 1236  
Db 520 ThrTyValProLysGluPheAsnAlaGluThrPheThrPheHisAlaAspIleCysThr 539  
Qy 1237 ATGTCCTCAA-----ATTGTTCCCGAGTGTATTATC 1266  
Db 540 LeuSerGluLysGluArgGlnIleLysLysGlnThrAlaLeuValGluLeuValLysHis 559  
Qy 1267 TATCCCGAA-----AAAGCATCTCTGATGAATAT----- 1296  
Db 560 LysProLysAlaThrLysGluGlnLeuLysAlaValMetAspAspPheAlaAlaPheVal 579  
Qy 1297 -----GACTACAGAAATCATCGACAAAAGAGACACCATATCTCTG 1338  
Db 580 GluLysCysCysLysAlaAspAspLysGluThrCysPheAlaGluGluGlyLysLysLeu 599  
Qy 1339 ACACCTGCTGTGATGCTCGGACTTGGAGTGCAGCAGGTGTAGGAACAGCAAGCTGCC 1398  
Db 600 ValAlaAlaSerGlnAlaLeuGlyLeuAlaPro-----ThrSerSerSerThr 616  
Qy 1399 -----CTGCTACGGGACCACAGCAGCTAGAAACAGGA 1431  
Db 617 LysLysThrGlnLeuGlnLeuGluHisLeuLeuLeuAspLeuGlnMetIleLeuAsnGly 636  
Qy 1432 CTTAGTAAC-----CTACATCGAATTGTAAACA----- 1458  
Db 637 IleAsnAsnTyLysAsnProLysLeuThrArgMetLeuThrPheLysPheTyMetPro 656  
Qy 1459 -----GAAGATCTCAAGCCCTAGAAAATCTGTCACTGTAACCTGGAG 1500  
Db 657 LysLysAlaThrGluLeuLysHisLeuGlnCysLeuGluGluGluLeuLysProLeuGlu 676  
Qy 1501 GAA-----TCCTTAACCTCC 1515  
Db 677 GluValLeuAsnLeuAlaGlnSerLysAsnPheHisLeuArgProArgAspLeuIleSer 696  
Qy 1516 TTATCTGAAGTAGTCTACAGAAATAGAGAGGGTTAGATTTATTATTCTTAAAAAGAGGA 1575  
Db 697 AsnIleAsnValIleValLeuGluLeuLysGlySerGluThrThrPheMet----- 713  
Qy 1576 GGATTATGTGTAGCCTTGAAGGAGGAATGCTGTTTTTTATGTGGATCATTCAGGGGCCATC 1635  
Db 714 -----CysGluTyAlaAspGluThrAlaThrIle 723  
Qy 1636 AGAGACTCCATGAACAAG 1653  
Db 724 ValGluPheLeuAsnArg 729  
RESULT 14  
US-11-175-690-534  
; Sequence 534, Application US/11175690  
; Publication No. US20060014254A1  
; GENERAL INFORMATION:  
; APPLICANT: Haseltine et al.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF605  
; CURRENT APPLICATION NUMBER: US/11/175,690  
; CURRENT FILING DATE: 2005-07-07  
; PRIOR APPLICATION NUMBER: PCT/US04/001369  
; PRIOR FILING DATE: 2004-01-20  
; PRIOR APPLICATION NUMBER: US 60/441,305  
; PRIOR FILING DATE: 2003-01-22  
; PRIOR APPLICATION NUMBER: US 60/453,201  
; PRIOR FILING DATE: 2003-03-11  
; PRIOR APPLICATION NUMBER: US 60/467,222  
; PRIOR FILING DATE: 2003-05-02  
; PRIOR APPLICATION NUMBER: US 60/472,816  
; PRIOR FILING DATE: 2003-05-23  
; PRIOR APPLICATION NUMBER: US 60/476,267  
; PRIOR FILING DATE: 2003-06-06  
; PRIOR APPLICATION NUMBER: US 60/505,172  
; PRIOR FILING DATE: 2003-09-24

```
; PRIOR APPLICATION NUMBER: US 60/506,746
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 534
; LENGTH: 742
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-175-690-534

Alignment Scores:
Pred. No.: 0.424 Length: 742
Score: 96.00 Matches: 64
Percent Similarity: 32.0% Conservative: 34
Best Local Similarity: 20.9% Mismatches: 88
Query Match: 2.8% Indels: 120
DB: 7 Gaps: 14

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-11-175-690-534 (1-742)
QY 1030 ACCCTTACTGAGTTTCTGGAAGGCACTGCATAGAAAGGTTCCCCATCCACCAA 1089
Db 446 ThrLeuValGluValSerArgAsn-----LeuGlyLysVal-----GlySer 459
QY 1090 CACCTTTGTAAACCACTGAAGCCTTAAATCAAACTCTGAGAGTCAATATCTG----- 1143
Db 460 LysCysCysLysHisProGluAlaLysArgMetProCysAlaGluAspTyrLeuSerVal 479
QY 1144 -----GTACCTGGTTATGACAGTGTGGGCA 1170
Db 480 ValLeuAsnGlnLeuCysValLeuHisGluLysThrProValSerAspArgValThrLys 499
QY 1171 TGTAAATACT-----GGATTAAACCCCTTGTGTTCACCTTGTTTAAACCAA 1218
Db 500 CysCysThrGluSerLeuValAsnArgProCysPheSerAlaLeuGluValAspGlu 519
QY 1219 ACT-----AAAGATTTTGCATT 1236
Db 520 ThrTyrValProLysGluPheAsnAlaGluThrPheThrPheHisAlaAspIleCysThr 539
QY 1237 ATGCTCCAA-----ATTGTTCCCGAGTGTATTAC 1266
Db 540 LeuSerGluLysGluArgGlnIleLysLysGlnThrAlaLeuValGluValLysHis 559
QY 1267 TATCCCGAA-----AAAGCAATCCTTGATGAATAT----- 1296
Db 560 LysProLysAlaThrLysGluLysAlaValMetAspAspPheAlaAlaPheVal 579
QY 1297 -----GACTACAGAAATCATCGACAAAGAGAGAACCCATATCTCTG 1338
Db 580 GluLysCysCysLysAlaAspAspLysGluThrCysPheAlaGluGluGlyLysLysLeu 599
QY 1339 ACATTCTGTGATCTCGGCTTGGAGTGGCAGCAGGTGTAGAAACAGGACAGCTGCC 1398
Db 600 ValAlaAlaSerGlnAlaLeuGlyLeuAlaPro-----ThrSerSerSerThr 616
QY 1399 -----CTGTCTACCGGACCCACAGCAGCTAGAAACAGGA 1431
Db 617 LysLysThrGlnLeuGlnLeuGluHisLeuLeuAspLeuGlnMetIleLeuAsnGly 636
QY 1432 CTATAGTAAC-----CTACATCGAATTGTAACA----- 1458
Db 637 IleAsnAsnTyrLysAsnProLysLeuThrArgMetLeuThrPheLysPheTyrMetPro 656
QY 1459 -----GAAGATCTCCAGCCCTAGAAAATCTGCAGTAACCTGGAG 1500
Db 657 LysLysAlaThrGluLeuLysHisLeuGlnCysLeuGluGluLysProLeuGlu 676
QY 1501 GAA-----TCCCTAACTCC 1515
Db 677 GluValLeuAsnLeuAlaGlnSerLysAsnPheHisLeuArgProArgAspLeuIleSer 696
QY 1516 TTATCTGAAGTAGTCTTACAGAATAGAGAGGGTTAGATTATTATTCTTAAAGAGGA 1575

; PRIOR APPLICATION NUMBER: US 60/506,746
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 534
; LENGTH: 742
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-175-690-534

Alignment Scores:
Pred. No.: 0.424 Length: 742
Score: 96.00 Matches: 64
Percent Similarity: 32.0% Conservative: 34
Best Local Similarity: 20.9% Mismatches: 88
Query Match: 2.8% Indels: 120
DB: 7 Gaps: 14

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-11-175-690-534 (1-742)
QY 1030 ACCCTTACTGAGTTTCTGGAAGGCACTGCATAGAAAGGTTCCCCATCCACCAA 1089
Db 446 ThrLeuValGluValSerArgAsn-----LeuGlyLysVal-----GlySer 459
QY 1090 CACCTTTGTAAACCACTGAAGCCTTAAATCAAACTCTGAGAGTCAATATCTG----- 1143
Db 460 LysCysCysLysHisProGluAlaLysArgMetProCysAlaGluAspTyrLeuSerVal 479
QY 1144 -----GTACCTGGTTATGACAGTGTGGGCA 1170
Db 480 ValLeuAsnGlnLeuCysValLeuHisGluLysThrProValSerAspArgValThrLys 499
QY 1171 TGTAAATACT-----GGATTAAACCCCTTGTGTTCACCTTGTTTAAACCAA 1218
Db 500 CysCysThrGluSerLeuValAsnArgProCysPheSerAlaLeuGluValAspGlu 519
QY 1219 ACT-----AAAGATTTTGCATT 1236
Db 520 ThrTyrValProLysGluPheAsnAlaGluThrPheThrPheHisAlaAspIleCysThr 539
QY 1237 ATGCTCCAA-----ATTGTTCCCGAGTGTATTAC 1266
Db 540 LeuSerGluLysGluArgGlnIleLysLysGlnThrAlaLeuValGluValLysHis 559
QY 1267 TATCCCGAA-----AAAGCAATCCTTGATGAATAT----- 1296
Db 560 LysProLysAlaThrLysGluLysAlaValMetAspAspPheAlaAlaPheVal 579
QY 1297 -----GACTACAGAAATCATCGACAAAGAGAGAACCCATATCTCTG 1338
Db 580 GluLysCysCysLysAlaAspAspLysGluThrCysPheAlaGluGluGlyLysLysLeu 599
QY 1339 ACATTCTGTGATCTCGGCTTGGAGTGGCAGCAGGTGTAGAAACAGGACAGCTGCC 1398
Db 600 ValAlaAlaSerGlnAlaLeuGlyLeuAlaPro-----ThrSerSerSerThr 616
QY 1399 -----CTGTCTACCGGACCCACAGCAGCTAGAAACAGGA 1431
Db 617 LysLysThrGlnLeuGlnLeuGluHisLeuLeuAspLeuGlnMetIleLeuAsnGly 636
QY 1432 CTATAGTAAC-----CTACATCGAATTGTAACA----- 1458
Db 637 IleAsnAsnTyrLysAsnProLysLeuThrArgMetLeuThrPheLysPheTyrMetPro 656
QY 1459 -----GAAGATCTCCAGCCCTAGAAAATCTGCAGTAACCTGGAG 1500
Db 657 LysLysAlaThrGluLeuLysHisLeuGlnCysLeuGluGluLysProLeuGlu 676
QY 1501 GAA-----TCCCTAACTCC 1515
Db 677 GluValLeuAsnLeuAlaGlnSerLysAsnPheHisLeuArgProArgAspLeuIleSer 696
QY 1516 TTATCTGAAGTAGTCTTACAGAATAGAGAGGGTTAGATTATTATTCTTAAAGAGGA 1575

Db 697 AsnIleAsnValIleValLeuGluLysGlySerGluThrThrPheMet----- 713
QY 1576 GGATTATGTGTAGCTTGAAGGAGGAATGCTGTTTTTATGTGGATCATTTCAGGGGCCATC 1635
Db 714 -----CysGluTyrAlaAspGluThrAlaThrIle 723
QY 1636 AGAGACTCCATGAACAAG 1653
Db 724 ValGluPheLeuAsnArg 729

RESULT 15
US-11-175-690-546
; Sequence 546, Application US/11175690
; Publication No. US20060014254A1
; GENERAL INFORMATION:
; APPLICANT: Haseltine et al.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF605
; CURRENT APPLICATION NUMBER: US/11/175,690
; CURRENT FILING DATE: 2005-07-07
; PRIOR APPLICATION NUMBER: PCT/US04/001369
; PRIOR FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: US 60/441,305
; PRIOR FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: US 60/453,201
; PRIOR FILING DATE: 2003-03-11
; PRIOR APPLICATION NUMBER: US 60/467,222
; PRIOR FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: US 60/472,816
; PRIOR FILING DATE: 2003-05-23
; PRIOR APPLICATION NUMBER: US 60/476,267
; PRIOR FILING DATE: 2003-06-06
; PRIOR APPLICATION NUMBER: US 60/505,172
; PRIOR FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: US 60/506,746
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 546
; LENGTH: 742
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-175-690-546

Alignment Scores:
Pred. No.: 0.424 Length: 742
Score: 96.00 Matches: 64
Percent Similarity: 32.0% Conservative: 34
Best Local Similarity: 20.9% Mismatches: 88
Query Match: 2.8% Indels: 120
DB: 7 Gaps: 14

US-10-723-552-3_COPY_5620_7533 (1-1914) x US-11-175-690-546 (1-742)
QY 1030 ACCCTTACTGAGTTTCTGGAAGGCACTGCATAGAAAGGTTCCCCATCCACCAA 1089
Db 446 ThrLeuValGluValSerArgAsn-----LeuGlyLysVal-----GlySer 459
QY 1090 CACCTTTGTAAACCACTGAAGCCTTAAATCAAACTCTGAGAGTCAATATCTG----- 1143
Db 460 LysCysCysLysHisProGluAlaLysArgMetProCysAlaGluAspTyrLeuSerVal 479
QY 1144 -----GTACCTGGTTATGACAGTGTGGGCA 1170
Db 480 ValLeuAsnGlnLeuCysValLeuHisGluLysThrProValSerAspArgValThrLys 499
QY 1171 TGTAAATACT-----GGATTAAACCCCTTGTGTTCACCTTGTTTAAACCAA 1218
Db 500 CysCysThrGluSerLeuValAsnArgProCysPheSerAlaLeuGluValAspGlu 519
QY 1219 ACT-----AAAGATTTTGCATT 1236
Db 520 ThrTyrValProLysGluPheAsnAlaGluThrPheThrPheHisAlaAspIleCysThr 539
QY 1237 ATGCTCCAA-----ATTGTTCCCGAGTGTATTAC 1266
Db 540 LeuSerGluLysGluArgGlnIleLysLysGlnThrAlaLeuValGluValLysHis 559
QY 1267 TATCCCGAA-----AAAGCAATCCTTGATGAATAT----- 1296
Db 560 LysProLysAlaThrLysGluLysAlaValMetAspAspPheAlaAlaPheVal 579
QY 1297 -----GACTACAGAAATCATCGACAAAGAGAGAACCCATATCTCTG 1338
Db 580 GluLysCysCysLysAlaAspAspLysGluThrCysPheAlaGluGluGlyLysLysLeu 599
QY 1339 ACATTCTGTGATCTCGGCTTGGAGTGGCAGCAGGTGTAGAAACAGGACAGCTGCC 1398
Db 600 ValAlaAlaSerGlnAlaLeuGlyLeuAlaPro-----ThrSerSerSerThr 616
QY 1399 -----CTGTCTACCGGACCCACAGCAGCTAGAAACAGGA 1431
Db 617 LysLysThrGlnLeuGlnLeuGluHisLeuLeuAspLeuGlnMetIleLeuAsnGly 636
QY 1432 CTATAGTAAC-----CTACATCGAATTGTAACA----- 1458
Db 637 IleAsnAsnTyrLysAsnProLysLeuThrArgMetLeuThrPheLysPheTyrMetPro 656
QY 1459 -----GAAGATCTCCAGCCCTAGAAAATCTGCAGTAACCTGGAG 1500
Db 657 LysLysAlaThrGluLeuLysHisLeuGlnCysLeuGluGluLysProLeuGlu 676
QY 1501 GAA-----TCCCTAACTCC 1515
Db 677 GluValLeuAsnLeuAlaGlnSerLysAsnPheHisLeuArgProArgAspLeuIleSer 696
QY 1516 TTATCTGAAGTAGTCTTACAGAATAGAGAGGGTTAGATTATTATTCTTAAAGAGGA 1575
```

